The Dissertation Committee for Lisa Anne Vasicek Certifies that this is the approved version of the following dissertation:

Enhanced Protein Characterization through Selective Derivatization and Electrospray Ionization Tandem Mass Spectrometry

Committee:

Jennifer Brodbelt, Supervisor

James Holcombe

Katherine Willets

Eric Anslyn

Hung-wen Liu

Enhanced Protein Characterization through Selective Derivatization and Electrospray Ionization Tandem Mass Spectrometry

by

Lisa Anne Vasicek, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2011

For my mother and her unwavering trust in my abilities.

Acknowledgements

The work compiled in this dissertation has been a collaborative effort in more ways than one. Whether through the guidance of a mentor, the advice of fellow group members, or the support and distractions provided when reactions refused to cooperate, all have been absolutely vital to the completion of this work.

I'd like to thank my advisor and mentor, Professor Jennifer Brodbelt, who provided an environment that allowed me to thrive in as a student and as a leader. Her patience, support, and drive have encouraged me to become a successful, independent scientist−one I hope she is proud to work alongside. To my undergraduate advisor, Professor Kenneth Roberts, thank you for believing I could complete this program and for guiding me to this "Wonderful World" in the first place.

I would like to thank my collaborators that provided essential components to the projects leading to this dissertation. Thanks to Professor Joshua Coon and Aaron Ledvina from the University of Wisconsin at Madison for the collaborative efforts to further photodissociation. I would like to thank the lab of Dr. Hung-wen Liu, Laura Mayberry and Dr. Karen Browning for providing me with interesting and biologically relevant samples to showcase my newly developed methods. I gratefully acknowledge Dr. Shagufta Shabbir, Colin Kubarych, Dr. Eric Anslyn, Jason Zbieg, Dr. Michael Krische, and John O'Brien for their guidance and assistance in all of my organic endeavors.

To my fellow Brodbelt group members, both past and present, I have been honored to struggle amongst you all. Our inability to separate life in lab from that outside has led to countless discussions that turned failing projects into fruitful ones. Special

thanks must be paid to Drs. Jeffrey Wilson, Haley Finley-Jones, Suncerae Smith, and James Madsen, for your friendships have motivated me and kept me going these five years. To the rest of my Austin/Belgian friends, thank you for providing a place of escape, for listening even though you don't always completely understand, and for making me laugh through it all.

Finally, I will be forever grateful to my family for their love and support through the past thirteen years of my life as a scientist. Thank you for fostering my creativity when I didn't know who I was, for challenging me to be my absolute best, and for always making me believe I can achieve the impossible goals I set for myself.

Enhanced Protein Characterization through Selective Derivatization and Electrospray Ionization Tandem Mass Spectrometry

Lisa Anne Vasicek, Ph.D.

The University of Texas at Austin, 2011

Supervisor: Jennifer S. Brodbelt

There continue to be great strides in the field of proteomics but as samples become more complex, the ability to increase sequence coverage and confidence in the identification becomes more important. Several methods of derivatization have been developed that can be used in combination with tandem mass spectrometry to identify and characterize proteins. Three types of activation, including infrared multiphoton dissociation, ultraviolet photodissociation, and electron transfer dissociation, are enhanced in this dissertation and compared to the conventional method of collisional induced dissociation (CID) to demonstrate the improved characterization of proteins.

A free amine reactive phosphate group was synthesized and used to modify the N-terminus of digested peptides. This phosphate group absorbs at the IR wavelength of 10.6 μm as well as the Vacuum-ultraviolet (VUV) due to an aromatic group allowing modified peptides to be dissociated by infrared multi-photon dissociation (IRMPD) or ultraviolet photodissociation (UVPD) whereas peptides without this chromophore are less responsive to IR or UV irradiation. The PD spectra for these modified peptides yield simplified MS/MS spectra due to the neutralization of all N-terminal product ions from the incorporation the negatively charged phosphate moiety. This is especially

advantageous for UVPD due to the great number of product ions produced due to the higher energy deposition of the UV photons. The MS/MS spectra also produce higher sequence coverage in comparison to CID of the modified or unmodified peptides due to more informative fragmentation pathways generated upon PD from secondary dissociation and an increased ion trapping mass range.

IRMPD is also implemented for the first time on an orbitrap mass spectrometer to achieve high resolution analysis of IR chromophore-derivatized samples as well as top-down analysis of unmodified proteins. High resolution/high mass accuracy analysis is extremely beneficial for characterization of complex samples due to the likelihood of false positives at lower resolutions/accuracies.

For electron transfer dissociation, precursor ions in higher charge states undergo more exothermic electron transfer and thus minimize non-dissociative charge reduction. In this dissertation, cysteine side chains are alkylated with a fixed charge to deliberately increase the charge states of peptides and improve electron transfer dissociation. ETD can also be used to study protein structure by derivatizing the intact structure with a hydrazone reagent. A hydrazone bond will be preferentially cleaved during ETD facilitating the recognition of any modified residues through a distinguishing ETD fragmentation spectrum.

Table of Contents

Chapter 1

Introduction

1.1 INTRODUCTION

There has been tremendous progress in the field of proteomics in the past decade, partly due to advances in mass spectrometric instrumentation and methodology. The successful elucidation of the primary sequence of many proteins/peptides through tandem mass spectrometry has pushed the development of many complementary dissociation methods that allow for more confident and faster identification of proteins and can provide information on amino acid mutations, post-translational modifications, and tertiary structure of the protein. Even still 75-85% of spectra in a typical MS/MS proteomic experiment remain unidentified.[1] This dissertation focuses on the development and optimization of selective derivatization to enhance tandem mass spectrometry methods for the analysis of proteins/peptides, specifically infrared multiphoton dissociation, ultraviolet photodissociation, and electron transfer dissociation.

1.2 TANDEM MASS SPECTROMETRY

Tandem mass spectrometry (MS/MS) refers to the selection/isolation and manipulation (energization or reaction) of a particular precursor ion of interest in which both the precursor and resulting product ions are characterized by their mass/charge ratios to gather a more detailed account of the precursor ion's original intact structure.[2] In the context of protein analysis the most common method is a bottom-up approach in which a

protein is enzymatically digested producing a set of peptides that are separated using liquid chromatography and analyzed by tandem mass spectrometry.[3] The resulting fragmentation patterns can be utilized to identify each peptide and in turn the protein. There are two current automated algorithm techniques that utilize the fragmentation spectra produced upon MS/MS analysis of peptides to identify the protein; in silico and de novo sequencing.[4] In silico algorithms use a large database of known protein sequences and compare hypothetical fragmentation spectra for the digested peptides to tandem MS spectra of real peptides to find the best match. De novo sequencing has greater impact for unknown protein sequences in that these proteins do not match up to any proteins in the in silico databases. This type of identification is done by searching adjacent peaks in a MS/MS spectrum for pairs of ions that have a m/z difference equivalent to the m/z of an amino acid. This process is repeated throughout the spectrum to sequence the entire peptide. However, this method can become complicated quickly with multiple ion types cluttering a spectrum or limited sequence coverage due to selective fragmentation along the backbone. Without a backbone cleavage between each amino acid, only partial sequence information can be determined, thus leading to inconclusive identification of the peptide/protein.

There are several methods of ion activation and dissociation in addition to recent advancements in ion-molecule and ion-ion reactions that allow for structural analysis of biomolecules.[2, 5, 6] Many of these activation methods are complementary to one another providing supplemental data when used cooperatively. No single method has proven to be universal for all mass spectrometers and all analyte systems so research into novel dissociation methods as well as ways to improve upon current protocols continues.[7-10]

1.2.1 Proteomic Fragmentation Nomenclature

Ion activation and dissociation of peptides/proteins typically produces a predictable set of cleavages along the amino acid backbone. Therefore a standardized nomenclature has been created to ease the identification of these ions where an alphabet letter indicates a specific fragment type, dictating the bond being broken, and a subscript number indicating the number of amino acid side chains contained in the ion following cleavage.[11] An example of this standard nomenclature is shown in **Figure 1.1**, where a four amino acid long peptide with side chains R_1-R_4, has been depicted with the most common dissociation sites marked. Tandem mass spectrometry typically results in cleavages along the peptide backbone as indicated by the lines cutting through a bond. For low energy dissociation techniques such as collision induced dissociation (CID) or infrared multiphoton dissociation (IRMPD), ergodic fragmentation is achieved through heating and the most thermally labile bonds, the backbone amides, are normally cleaved producing *b* and *y* ions, in which the *b*-type ion contains the N-terminus and the *y*-type ion contains the complementary C-terminus. An additional loss of CO from the *b* ion produces the associated *a* ion. Electron based dissociation methods such as electron capture dissociation (ECD)[12, 13] and electron transfer dissociation (ETD)[14] produce radical driven fragmentation of the N-C_α bond leading to *c*- and *z*- ion types. The homolytic backbone cleavage of the C_α-C bond produced upon yields *a*- and *x*- ion types in addition to some secondary side chain cleavages that yield *v*-, *w*-, or *d*-type ions depending on where the charge is located.[15, 16]

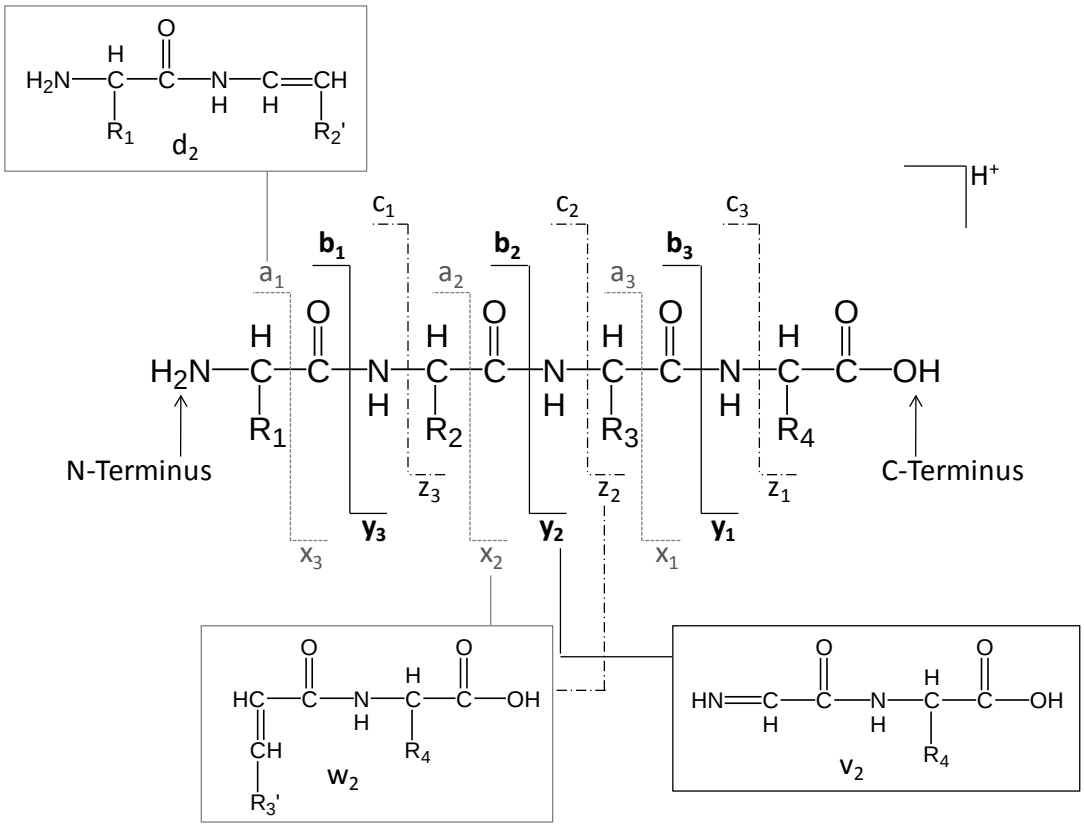

Figure 1.1 Peptide/Protein Fragmentation Nomenclature

1.2.2 Collision-Induced Dissociation (CID)

The most commonly used activation method for proteomic applications is collision-induced dissociation in which a selected precursor is activated through collisions.[7, 17, 18] The precursor ion is accelerated by application of a voltage to the mass spectrometer that results in energetic collisions with gas molecules, thus converting the increased kinetic energy of the ion into internal energy that leads to dissociation. The total energy available for the transfer from kinetic energy to internal energy depends on the ion's kinetic energy as well as the mass of the target molecule and precursor ion. In addition, the pressure of the collision cell is of great importance as at higher pressures the

probability of multiple collisions increases. To ensure multiple energetic collisions, most CID experiments are carried out at ~1 mTorr.[19] CID is the most established dissociation technique as it can be performed in a wide range of instruments, including time-of-flight (TOF), Fourier transform ion cyclotron resonance (FTICR), tandem quadupole and ion traps (IT), and is well-suited for multiple analyte types. However, CID is still not without its limitations. For example, in trapping instruments the rf potential must be ramped to increase the kinetic energy of the selected ion. This rf is also used to trap ions, causing a compromise between efficiently activating the precursor and retaining small mass-to-charge (m/z) fragment ions for detection.[20] This compromise is known as the low-mass cut-off (LMCO) and for most quadrupole ion trap instruments causes the loss of the lower fourth of the mass range in relation to the precursor m/z. In addition, as a resonant dissociation method, only the precursor ion is activated and the dominant dissociation pathway may be uninformative, such as the loss of small neutral molecule like water or NH_3. Several stages of MS/MS (commonly referred to as MS^n) can be used to overcome this problem, where a product ion is subjected to a second stage of activation, but these experiments can prove time consuming and result in reduced sensitivity. Lastly, as a collision-based method, CID may cause ion losses due to scattering as an ion's motion becomes destabilized from multiple collisions. For these reasons alternative activation methods are being developed as complementary techniques to CID and to alleviate some of the current limitations of CID alone.

1.2.3 Photodissociation

Photodissociation is an alternative activation method that relies upon exciting an ion population through exposure to photons produced from a laser. Successful dissociation is dependent on the ion population's ability to absorb the photons upon irradiation. Depending on the wavelength used, dissociation can occur following a single photon as seen with high energy UV photons (~3-10 eV) or hundreds of photons as with low energy IR photons (~0.1 eV). Low energy IR photons result in vibrational excitation similar to that of CID while excitation by a higher energy UV photon leads to electronic excitation providing very different dissociation pathways.

Photodissociation provides many solutions to the limitations discussed previously for CID. For trapping instruments, photodissociation is independent of the rf trapping voltage and therefore a broader mass range of product ions can be detected. In addition, as an absorption process, an ion's motion is not disrupted and ion scattering is greatly reduced. Unlike CID, primary product ions created upon photodissociation may simultaneously undergo dissociation as they are also being irradiated, so uninformative pathways can be surpassed due to secondary dissociation. This yields more informative fragmentation pathways that were only previously available through MS^n with CID.

The efficiency of photon absorption is essential to successful photodissociation experiments and can be dramatically influenced by a subtle shift in chemical structure such as a substituent or protonation. While this leads to the great advantage of photodissociation being tunable through varied irradiation times, laser power or laser wavelength, it also means a single wavelength may not be suitable for all analytes.

However, infrared multiphoton dissociation (IRMPD) is commonly implemented with a CO_2 laser at $\lambda = 10.6$ μm which can be absorbed by a variety of vibrational modes (C-C, C-N, C-O, P-O, etc), giving it the potential to have a widespread application base. IRMPD has been successful in FTICRs and trapping instruments.[21] In ion trapping instruments, the biggest disadvantage of IRMPD is the competition with collisional cooling by the bath gas at optimal trapping pressures (~1 mTorr). Most analytes can be excited by IRMPD but at the general operating pressure for ion traps the rate of collisional cooling is nearly equal that of the photon absorption rate, limiting dissociation unless a strong chromophore is present.[5, 22] There have been several methods developed to overcome this limitation of IRMPD, including heating the bath gas to decrease collisional cooling[23], applying a supplemental collisional activation to the precursor to increase the initial internal energy of precursor prior to irradiation[24], focusing the laser for higher photon energies[25, 26], lower the trap pressure so the bath gas is reduced during activation,[27, 28] and covalent chemical modification to incorporate a chromophore into the structure of the analyte of interest to increase the IR absorption cross-section.[29-33] Such methods will be discussed in this dissertation.

Ultraviolet photodissociation is gaining interest recently for the analysis of biomolecules. The large range of wavelengths available in the visible and UV range allow for the use of a diverse set of lasers that produce photon energies up to 100 times greater than those of IRMPD.[34, 35] In particular, aromatic amino acid side chains have naturally occurring electronic transitions that can be excited via an excimer laser at 157 nm (F_2 laser) or 193 nm (ArF laser) and the fourth harmonic of a Nd:YAG laser at 266 nm.[15, 35-39] At these wavelengths, the photons are energetic enough that a single photon is

enough to produce fragmentation, yielding a wide array of fragments including *a*-, and *x*-type ions and the unique side chain cleavages that allow for distinction between isobaric residues such as Leu and Iso.[35, 40] However, peptides without aromatic side chains do not display as intense absorption of UV photons. Increasing the aromaticity of these peptides through derivatization can lead to increased UVPD and is discussed in chapter 5.

1.2.4 Electron Transfer Dissociation

Electron-based activation methods are becoming a popular complementary technique to CID and have shown great potential in the field of proteomics. The two most promising are electron capture dissociation[41] and electron transfer dissociation.[14, 42] Electron capture (ECD) was developed in FTICR instruments in which a multiply protonated peptide is exposed to a stream of low energy electrons. The capture of an electron causes dissociation in a sequence independent manner. ECD requires a large population of low energy electrons that cannot be readily trapped in a radio frequency field, making it impossible in quadrupole ion traps.[13, 43] In order to implement a similar type of dissociation in other trapping instruments, ion/ion reactions are utilized. A reagent anion with low electron affinity is reacted with a multiply protonated peptide, and an electron is transferred from the anion to the peptide in an exothermic process creating an odd-electron species.[14] This activated, charge-reduced species will undergo subsequent fragmentation at the N-C_α bond, similar to ECD, producing *c*- and *z*-type fragment ions rather than the *y*- and *b*-type ions that are produced by thermal dissociation methods like CID (see **Figure 1.1**). This process has been termed electron transfer dissociation (ETD) and can be employed in both linear and quadrupole ion traps. As a nonergodic process,

there is little energy redistribution before dissociation, affording more random cleavages along the backbone and also allowing retention and identification of post translational modifications.[44, 45] All of these distinctions make ETD and ECD complementary to CID.

As with most methods, ETD also has some disadvantages as well. As electron transfer is necessary prior to dissociation, these methods are not suited for singly protonated precursors and are more efficient for higher multiply charged precursors (3+, 4+, 5+, etc.).[46, 47] For doubly protonated precursors, the charge reduced species does not undergo efficient dissociation without supplemental activation. There have been methods developed to remedy this latter problem such incorporation of additional charges into the structure or the application of a supplemental collision activation step to apply a small amount of energy to excite the charge reduced species triggering dissociation.[48-55] Another method to bypass this shortcoming through the incorporation of additional charges is discussed in chapter 4.

1.3 SELECTIVE DERIVATIZATION

The most common proteomics protocol for mass spectrometry involves a bottom-up approach for identification as shown in **Figure 1.2**, often involving complex mixtures of proteins from cells or tissues.[2, 56] A cell lysate releases thousands of proteins, lipids, DNA and biological materials. A group of the separated proteins can then be enzymatically digested and separated/analyzed by LC MS/MS. Subsequent protein identification is achieved using database searching of the MS/MS data recorded in an in silico or *de novo* manner.

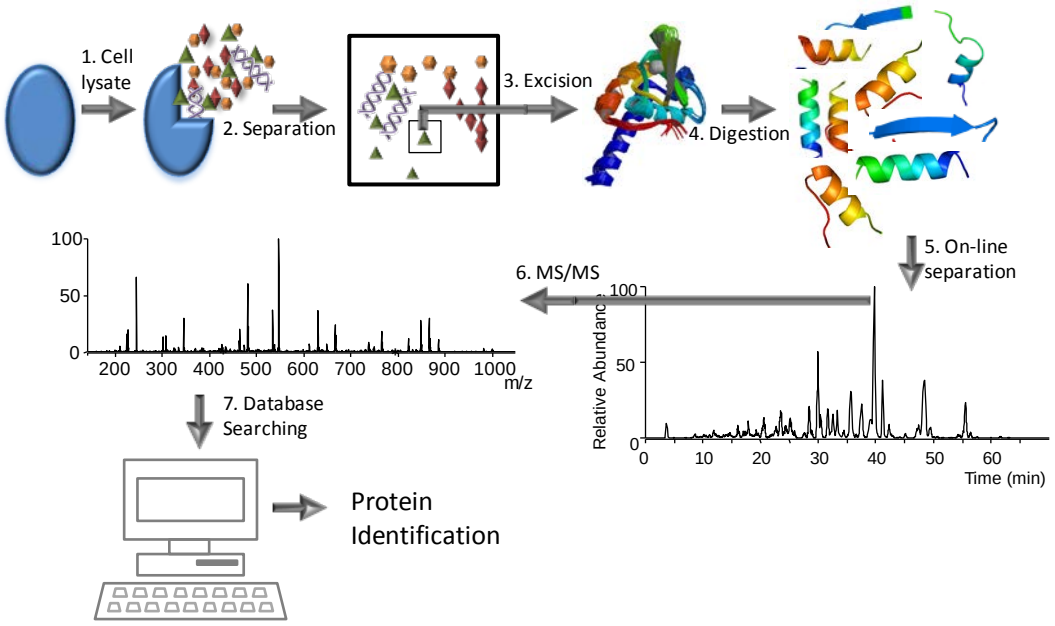

Figure 1.2 General Protocol for Bottom-up Mass Spectrometry; 1. Cell is lysed to release all biological material; 2. Biological material is separated; 3. Excision of a group or single protein; 4. Protein enzymatically digested; 5. Digested peptides are separated through LC/ESI-MS; 6. Individual peptides analyzed through MS/MS; 7. MS/MS spectra are submitted to a software algorithm for identification.

However, even a fraction of the original cell lysate may contain a large number of proteins, which when digested can increase the sample complexity by several orders of magnitude limiting the performance of current methods employed for protein identification. One method that can simplify the analysis of these complex samples is selective derivatization.[29, 57-63] Chemical derivatization allows a specific group of

analytes with the same functional group to be targeted limiting the potential candidates for protein identification. In combination with a selective affinity extraction method, chemical derivatization can be used to decrease the complexity of the sample. For example, cysteines are only present in 17% of the tryptic peptides produced from the human proteome but over 89% of proteins contain at least one cysteine residue.[60] If only cysteine peptides are enriched from a complex mixture for analysis, then a large amount of the proteome can be identified through the analysis of a small number of peptides. If a reaction is selective for only thiol groups, then cysteine residues can be selectively modified and enriched through separation techniques leading to increased detection sensitivity for the modified peptides.

In addition, selective derivatization can be utilized to incorporate a functional group that can enhance mass spectrometric analysis.[64, 65] Electrospray ionization is one of the most amenable ionization techniques for proteomic analysis because it is a soft ionization technique meaning that the parent molecule is ionized without causing fragmentation in the process. This is important for proteomic studies so that the molecular weight of the original protein can be easily established. The large number of basic residues present in proteins (Lys, Arg and His) make positive ionization of proteins a favorable process. However, additional charges incorporated through derivatization, whether they are positive or negative, can enhance both the initial ionization of the protein, increasing sensitivity, and the subsequent fragmentation of the protein for more confident identification of the protein through increased sequence coverage.[66] Dissociation can also be altered through derivatization.[67-70] As already mentioned, chromophores can be added into a protein structure that allows for the protein to readily

absorb photons of a certain wavelength, thus enhancing photodissociation or the creation of a completely new set of fragment ions through ultraviolet photodissociation.

Selective covalent labeling has also been used to study the tertiary structure of proteins and their interactions with other proteins through a measure of reactivity.[58, 71-74] In addition to the affinity of the reagent for the amino acid, an amino acid's reactivity is based on the accessibility of its side chain to the reagent. Upon reaction, the extent of reaction for each possible amino acid can be compared and related to its position in the tertiary structure of the protein in terms of relative surface accessibility.[71] The surface accessibility information can be used to place constraints on the positions of the reacted amino acids and to predict the three-dimensional structure of a protein. This technique can also be used to investigate protein complexes if upon a protein-protein/protein-ligand interaction a conformational change occurs affecting the extent of accessibility of certain amino acids and changing their reactivity with the labeling reagent. If a change in accessibility occurs upon interaction (whether it be an increase or decrease), it is indicative of the involvement of that amino acid in a structural change and/or the location of that amino acid at the binding site.[71]

Although monitoring the reactivity of an intact protein by MS is quick and simple, the amount of information gleemed is not as detailed as that produced upon analysis of the enzymatically digested peptides produced following modification. Therefore the typical protocol for studying the structure of a protein using MS is to react the labeling reagent with the intact protein in solution followed by a bottom-up MS/MS approach as shown in **Figure 1.3**.

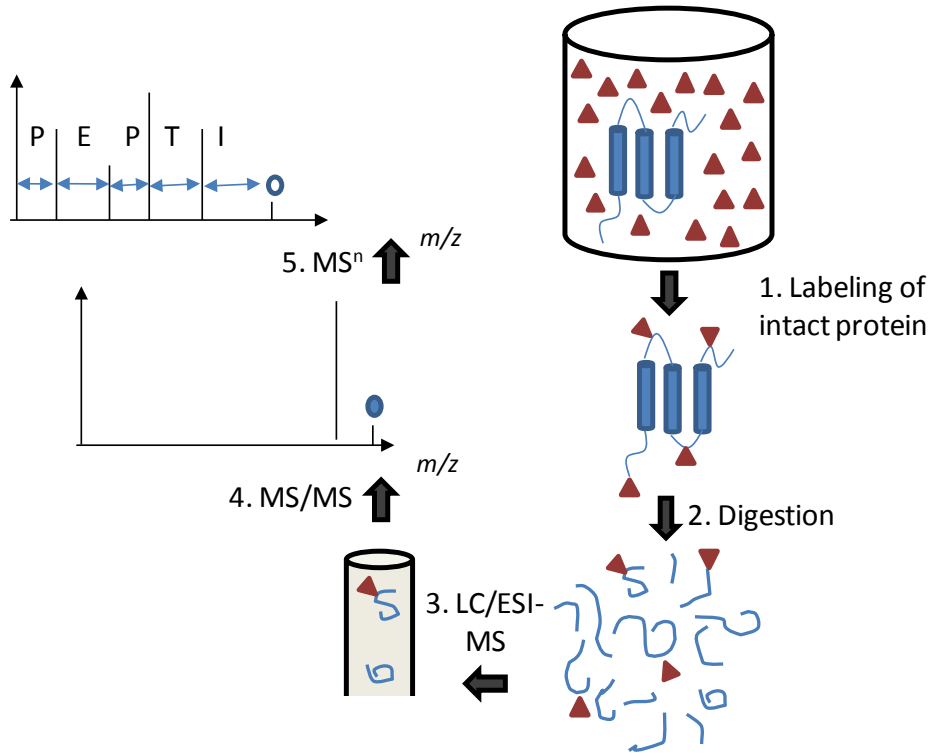

Figure 1.3 General Protocol for Surface Accessibility Study with MS; 1. Intact protein is labeled in solution and desalted; 2. Modified protein is enzymatically digested; 3. Digested peptides are separated and analyzed with LC/ESI-MS; 4. Modified peptides are identified by a characteristic MS/MS spectrum; 5. Modified peptides identified using MS^n.

Enzymatic digestion adds complexity to a sample and the introduction of possible modification sites will increase the intricacy of data analysis even more. Since identification of the modification sites is the key outcome in this type of analysis, it would be beneficial to be able to quickly separate all labeled peptides from those unmodified to eliminate a large part of the peptide population needing to be analyzed. One strategy to quickly pinpoint the modified peptides is through the use a characteristic MS/MS pattern. The Reid group has demonstrated this concept using sulfonium ions and

CID to quickly identify which peptides have been modified from the complex digest mixture and then utilize a MS^n stage to further interrogate the peptide.[64, 75-78]

1.4 OVERVIEW OF CHAPTERS

The focus of the research discussed in this dissertation is the development of derivatization methods that will selectively modify peptides or proteins with the aims of improving the sequencing or structural analysis. Derivatization methods can be utilized to attach chromophores, additional charges or mass tags that can be used to improve dissociation efficiency, increase ionization efficiency, or provide the means to improve separation from other components in a complex mixture.

Photodissociation offers key advantages to improve dissociation spectra of biomolecules. An IR chromophore with reactivity that is selective for free amines (Lysines and the N-terminus) was synthesized to increase the photoabsorptivity of tryptic peptides (Chapter 3). Upon tryptic digestion, most peptides will contain a lysine in the sequence, thus providing a convenient site of modification for each peptide allowing efficient dissociation upon IR irradiation. The modification of lysines with a phosphate chromophore can extend the use of IRMPD to cover a larger population of analytes for bottom-up proteomics.

High resolution instruments have expanded the potential for mass spectrometric analysis of large and small molecules. The increased mass accuracy allows for confident assignment of chemical formula for small molecules and a deconvoluted spectrum of higher charged species providing a facile method for determining its molecular weight. Photodissociation has long since been commercially available in expensive high

resolution instruments like the FTICR. But until the emergence of the Orbitrap providing high resolution at a lower price tag, the possibility of obtaining photodissociation with high mass accuracy and resolution to the masses has not been possible. Through the incorporation of an IR laser into the higher collision dissociation (HCD) cell, IRMPD in an orbitrap can be utilized to study chromophore enhanced or supercharged peptides for bottom-up proteomics and highly charged intact proteins for top-down analysis (Chapter 4).

Ultraviolet photodissociation provides higher energy deposition than IRMPD, promoting high energy fragmentation pathways. Aromatic residues have proven to be a natural UV chromophore promoting highly efficient UVPD spectra. However, peptides without aromatic residues provide minimal fragmentation that is not conducive to sequencing peptides. Through the attachment of an aromatic substituent, UVPD can be enhanced providing high energy fragmentation spectra for all peptides (Chapter 5). In addition a negative charge fixed on to the modification site can limit the type of fragment ions produced, leading to simplified fragmentation spectra. The same phosphate aromatic chromophore used to improve IRMPD in Chapter 3 can be used to enhance UVPD through the incorporation of an aromatic chromophore while the negatively charged phosphate substituent will neutralize many of the singly charged fragment ions that contain the modification.

Electron transfer dissociation has been shown to be a complementary dissociation technique to collision induced dissociation. CID is better suited for low charge states (singly and doubly charged species) while ETD of low charged species leads to only minor backbone fragmentation and primarily a charged reduced species that yields no

sequence information. Through the addition of a fixed charge to all reduced Cys residue, the ETD efficiency of all Cys peptides can be increased (Chapter 6). Ions in higher charge states promote more efficient formation of the diagnostic *c*- and *z*-type ions, leading to better sensitivity and increased confidence in peptide and protein identifications.

Finally, structural analysis of proteins is of great importance as it can give insight into the protein's function. Mass spectrometry in combination with surface accessibility methods yields low resolution structural information about a protein's surface in which selective derivatization of the native protein is used to mark key amino acids on the surface of a protein to distinguish them from the interior of the protein structure. After derivatization, the protein is typically subjected to proteolytic digestion and LCMS/MS analysis in which the peptides containing the surface accessibility tag are present amidst a large number of unmodified peptides. With a complex mixture of peptides containing both modified and unmodified ones, a method that allows for quick distinction between the two is desirable. It has been shown that a N-N hydrazone bond is preferentially cleaved by ETD. A novel Lys selective modification reagent that incorporates a N-N hydrazone bond onto all free amines was synthesized (Chapter 8). Due to the preferential cleavage of the N-N bond upon ETD, any modified species can be easily distinguished from unmodified species by monitoring the neutral loss created upon N-N cleavage.

1.5 REFERENCES

(1) Frank, A. M.; Monroe, M. E.; Shah, A. R.; Carver, J. J.; Bandeira, N.; Moore, R. J.; Anderson, G. A.; Smith, R. D.; Pevzner, P. A. *Nat. Methods*, *8*, 587-591.

(2) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233-6237.

(3) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806-812.

(4) Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *J. Comput. Biol.* **1999**, *6*, 327-342.

(5) McLuckey, S. A.; Goeringer, D. E. *J. Mass Spectrom.* **1997**, *32*, 461-474.

(6) McLuckey, S. A. *J. Mass Spectrom.* **1997**, *16*, 429-436.

(7) Wells, J. M.; McLuckey, S. A. *Methods Enzymol.* **2005**, *402*, 148-185.

(8) Laskin, J.; Futrell, J. H. *Mass Spectrom. Rev.* **2005**, *24*, 135-167.

(9) Dongre, A. R.; Somogyi, A.; Wysocki, V. H. *Journal of Mass Spectrometry* **1996**, *31*, 339-350.

(10) Chaurand, P.; Luetzenkirchen, F.; Spengler, B. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 91-103.

(11) Roepstorff, P. *Biomedical Mass Spectrometry* **1984**, *11*, 601.

(12) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.

(13) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563-573.

(14) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528-9533.

(15) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1657-1665.

(16) Zhang, L.; Reilly, J. P. *Anal. Chem.* **2009**, *81*, 7829-7838.

(17) McLafferty, F. W.; Venkataraghavan, R.; Irving, P. *Biochem. Biophys. Res. Commun.* **1970**, *39*, 274-278.

(18) Papayannopoulos, I. A. *Mass Spectrom. Rev.* **1995**, *14*, 49-73.

(19) Stafford, G. C., Jr.; Kelley, P. E.; Syka, J. E. P.; Reynolds, W. E.; Todd, J. F. J. *Int. J. Mass Spectrom. Ion Processes* **1984**, *60*, 85-98.

(20) Sleno, L.; Volmer, D. A. *Journal of Mass Spectrometry* **2004**, *39*, 1091-1112.

(21) Van der Hart, W. J. *Mass Spectrom. Rev.* **1989**, *8*, 237-268.

(22) Black, D. M.; Payne, A. H.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 932-938.

(23) Payne, A. H.; Glish, G. L. *Anal. Chem.* **2001**, *73*, 3542-3548.

(24) Hashimoto, Y.; Hasegawa, H.; Yoshinari, K.; Waki, I. *Anal. Chem.* **2003**, *75*, 420-425.

(25) Drader, J. J.; Hannis, J. C.; Hofstadler, S. A. *Anal. Chem.* **2003**, *75*, 3669-3674.

(26) Newsome, G. A.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1127-1131.

(27) Hashimoto, Y.; Hasegawa, H.; Waki, I. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2255-2259.

(28) Boue, S. M.; Stephenson, J. L., Jr.; Yost, R. A. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1391-1397.

(29) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80*, 4807-4819.

(30) Wilson, J. J.; Brodbelt, J. S. *Anal. Chem.* **2006**, *78*, 6855-6862.

(31) Pikulski, M.; Wilson, J. J.; Aguilar, A.; Brodbelt, J. S. *Anal. Chem.* **2006**, *78*, 8512-8517.

(32) Pikulski, M.; Hargrove, A.; Shabbir, S. H.; Anslyn, E. V.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2094-2106.

(33) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 377-384.

(34) Ly, T.; Julian, R. R. *Angew. Chem., Int. Ed.* **2009**, *48*, 7130-7137.

(35) Reilly, J. P. *Mass Spectrom. Rev.* **2009**, *28*, 425-447.

(36) Oh, J. Y.; Moon, J. H.; Kim, M. S. *Journal of Mass Spectrometry* **2005**, *40*, 899-907.

(37) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.* **2010**, *9*, 4205-4214.

(38) Oh, J. Y.; Moon, J. H.; Kim, M. S. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2706-2712.

(39) Oh, J. Y.; Moon, J. H.; Lee, Y. H.; Hyung, S.-W.; Lee, S.-W.; Kim, M. S. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1283-1288.

(40) Thompson, M. S.; Cui, W.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1439-1452.

(41) Zubarev, R. A. *Curr. Opin. Biotechnol.* **2003**, *15*, 12-16.

(42) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 880-882.

(43) Breuker, K.; Oh, H.; Lin, C.; Carpenter, B. K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14011-14016.

(44) McAlister, G. C.; Berggren, W. T.; Griep-Raming, J.; Horning, S.; Makarov, A.; Phanstiel, D.; Stafford, G.; Swaney, D. L.; Syka, J. E. P.; Zabrouskov, V.; Coon, J. J. *J. Proteome Res.* **2008**, *7*, 3127-3136.

(45) Shi, S. D. H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal. Chem.* **2001**, *73*, 19-22.

(46) Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *Anal. Chem.* **2005**, *77*, 5662-5669.

(47) Iavarone, A. T.; Paech, K.; Williams, E. R. *Anal. Chem.* **2004**, *76*, 2231-2238.

(48) Ledvina, A. R.; Beauchene, N. A.; McAlister, G. C.; Syka, J. E. P.; Schwartz, J. C.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2010**, *82*, 10068-10074.

(49) Li, X.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1514-1526.

(50) Chamot-Rooke, J.; van der Rest, G.; Dalleu, A.; Bay, S.; Lemoine, J. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1405-1413.

(51) Gunawardena, H. P.; Gorenstein, L.; Erickson, D. E.; Xia, Y.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2007**, *265*, 130-138.

(52) Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 6407-6420.

(53) Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W. *J. Am. Chem. Soc.* **2001**, *123*, 9792-9799.

(54) Horn, D. M.; Ge, Y.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 4778-4784.

(55) Ueberheide, B. M.; Fenyo, D.; Alewood, P. F.; Chait, B. T. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 1-6, 6 pp.

(56) Chait, B. T. *Science* **2006**, *314*, 65-66.

(57) Brancia, F. L.; Butt, A.; Beynon, R. J.; Hubbard, S. J.; Gaskell, S. J.; Oliver, S. G. *Electrophoresis* **2001**, *22*, 552-559.

(58) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2005**, *77*, 7274-7281.

(59) Vasicek, L.; Brodbelt, J. S. *Anal. Chem.* **2009**, *81*, 7876-7884.

(60) Regnier, F. E.; Riggs, L.; Zhang, R.; Xiong, L.; Liu, P.; Chakraborty, A.; Seeley, E.; Sioma, C.; Thompson, R. A. *J. Mass Spectrom.* **2002**, *37*, 133-145.

(61) Steen, H.; Kuester, B.; Fernandez, M.; Pandey, A.; Mann, M. *Anal. Chem.* **2001**, *73*, 1440-1448.

(62) Molloy, M. P.; Andrews, P. C. *Anal. Chem.* **2001**, *73*, 5387-5394.

(63) Leitner, A.; Linder, W. *Proteomics* **2006**, *6*, 5418-5434.

(64) Reid, G. E.; Roberts, K. D.; Simpson, R. J.; O'Hair, R. A. J. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1131-1150.

(65) Knapp, D. R. *Methods Enzymol.* **1990**, *193*, 314-329.

(66) Roth, K. D.; Huang, Z. H.; Sadagopan, N.; Watson, J. T. *Mass Spectrom. Rev.* **1998**, *17*, 255-274.

(67) Anderegg, R. J. *Mass Spectrom. Rev.* **1988**, *7*, 395-424.

(68) Sadagopan, N.; Watson, J. T. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 399-409.

(69) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348-2356.

(70) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Anal. Chem.* **2003**, *75*, 156A-165A.

(71) Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785-815.

(72) Downard, K. M. *Proteomics* **2006**, *6*, 5374-5384.

(73) Maleknia, S. D.; Downard, K. *Mass Spectrom. Rev.* **2001**, *20*, 388-401.

(74) Sinz, A. *Mass Spectrom. Rev.* **2006**, *25*, 663-682.

(75) Zhou, X.; Lu, Y.; Wang, W.; Borhan, B.; Reid, G. E. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1339-1351.

(76) Sierakowski, J.; Amunugama, M.; Roberts, K. D.; Reid, G. E. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1230-1238.

(77) Roberts, K. D.; Reid, G. E. *J. Mass Spectrom.* **2007**, *42*, 187-198.

(78) Froelich, J. M.; Kaplinghat, S.; Reid, G. E. *Eur. J. Mass Spectrom.* **2008**, *14*, 219-229.

Chapter 2

Experimental

2.1 Mass Spectrometry

Two types of mass spectrometers were utilized in the work discussed in this dissertation, with all using electrospray ionization (ESI) for generation of ions. The majority of the work was conducted on a ThermoFisher Scientific LTQ XL linear ion trap (San Jose, CA) modified for photodissociation and electron transfer dissociation. The geometry of a linear ion trap allows for an increased storage capacity of ions offering improved sensitivity in comparison to 3D quadrupoles. A linear ion trap consists of four rods arranged in a quadrupole and split into three sections with apertures on the front and back ends of the quadrupole. RF voltages are applied on the front and back sections of the quadrupole to confine the ions radially and can be raised and lowered to bring ions into the trap and move them through the trap. The ions are ejected radially through slits in the rods as the RF voltages are raised according to their m/z ratio with the smallest ions being moved first followed by the larger ions. Two electron multiplier detectors are positioned on either side of the ion trap to detect the ions as they are ejected from the ion trap. A Bruker MaXis time-of-flight mass spectrometer was also used for some experiments in this dissertation. A time-of-flight mass spectrometer uses differences in travel time through a field free drift tube to separate ions based on their mass-to-charge ratio. A strong electric field accelerates a population of ions from the same place at the same time into the tube. The velocity of each ion is related to its mass-to-charge ratio so

lighter ions will have a higher velocity than heavier ions and will reach the detector at the end of the drift tube at a faster rate.

ESI is a soft ionization source meaning that it produces ions without causing fragmentation so the mass of the initial analyte can be easily determined. Ions are produced by applying a potential to a heated capillary through which the analyte and a nitrogen gas flows, both nebulizing and desolvating the sample to produce multiply charged analyte ions. This heated capillary is maintained at 180 °C for all experiments to promote desolvation. Protein and peptide samples were typically run at 10 μM in 49.5:49.4:1 H_2O/MeOH/acetic acid (v/v/v).

2.2 INFRARED MULTIPHOTON DISSOCIATION (IRMPD)

IRMPD experiments were performed on a modified LTQ and modified Orbitrap mass spectrometers. The LTQ was modified through the introduction of a CF viewport flange with a ZnSe window with an anti-reflective argon coating in place of the vacuum manifold back flange to allow introduction of the 10.6 μm radiation. A 48-5 Synrad 50 W CO2 continuous wave laser (Mukilteo, WA) was triggered during ion activation using a TTL signal from pin 14 of the J1 connector and controlled through instrument software. The laser beam was passed through a 1.8 mm aperture that was mounted to the back flange and a 2 mm aperture of the exit lens prior to reaching the ion cloud in the linear ion trap. For the orbitrap, IRMPD was performed in the HCD cell of a modified hybrid mass spectrometer comprising both linear ion trap and orbitrap mass analyzers. The HCD manifold was modified *via* the addition of a ZnSe window co-axial with the HCD cell and opening an aperture (d = ~2.54 mm) in its end plate electrode, allowing the irradiation of the HCD trapping region with IR

photons. Additionally, a nitrogen gas line was installed in the C-trap and the top and bottom of the C-trap were fitted with ceramic plates to decrease gas conductance out of the C-trap, reducing the gas flow needed to achieve desired trapping efficiencies in the C-trap while minimizing the N_2 partial pressure in the QLT (quadrupole linear trap). Taken as a whole, these modifications allowed substantially independent control of nitrogen gas flow to both the C-trap and HCD cell. A 100 W Synrad CO_2 laser was directed in the HCD cell using two optic mirrors mounted behind the HCD cell. All experiments were conducted using 50 W power with irradiation times varying between 1 and 200 ms. A low activation q_z-value ($\leq$0.2) was used to minimize the low mass cut-off.

2.3 ULTRAVIOLET PHOTODISSOCIATION

UVPD was performed on a a ThermoFinnigan LTQ XL linear ion trap mass spectrometer using a Coherent Excistar XS 500 ArF excimer laser (Munich, Germany) with a repetition rate of 500 Hz and a pulse width of 5 ns. A mixture of argon and a small amount of fluorine was used as the active laser medium. Most UVPD irradiation was done with a single laser pulse triggered through the LTQ software as described above for IRMPD with the lowest possible activation setting on the LTQ software, 0.03 ms. All experiments were conducted using an energy of 1 to 6 mJ and was controlled with the Excistar software. The laser beam was introduced through the a CF viewport flange in place of the vacuum manifold back flange with a CaF_2 window with an anti-reflective 193 nm coating to allow complete transmission of the 193 nm laser. The activation q_z-value was set to 0.1 to minimize the low mass cut-off.

2.4 ELECTRON TRANSFER DISSOCIATION

ETD experiments were performed on the LTQ XL two dimensional linear ion trap and MaXis time-of-flight mass spectrometers. Fluoranthene anions were used as the reagent anion in both instruments. The reagent anion population was varied between 1 x 10^5 to 1 x 10^6 ions and generated by chemical ionization prior to being directed into the trap. All ion-ion reaction times varied between 100 and 250 ms.

2.5 CHEMICALS AND REAGENTS

All chemicals were purchased from Sigma Aldrich (St. Louis, MO) except for the following peptides which were obtained from BACHEM (King of Prussia, PA): FQVVCG, HCKFWW, CDPGYIGSR, NRCSQGSCWN, somatostatin (AGCKNFFWKTFTSC), LQVQLSIR, YGGFLK, FSWGAEGQR, bradykinin (RPPGFSPFR), TGF-alpha (CHSGYVGVRC), bactenecin (RLCRIVVIRVCR), amyloid protein 1-16 (DAEFRHDSGYQVHHQK), ASHLGLAR, YRPPGFSPFR, GNHWAVGHLM, MEHFRWG, and RQpSVELHSPQSLPR. HSDAVFTDNYTR and fibrinopeptide a (ADSGEGDFLAEGGGVR) was obtained from the American Peptide Co. (Sunnyvale, CA). Angiotensin I(DRVYIHPFHL), ubiquitin (bovine red blood cells), cytochrome c (equine heart), and BSA were purchased from Sigma (St. Louis, MO). 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC), succinimidyl 4-formylbenzoate (SFB), and immobilized trypsin beads were obtained from Pierce Biotechnology (Rockford, IL). GluC was purchased from New England BioLabs Inc. (Ipswich, MA). Water, acetonitrile, acetic acid and formic acid were all purchased from Fisher Scientific (Fairlawn, NJ).

2.6 CHEMICAL SYNTHESIS

For the work described in this dissertation, a number of custom compounds were designed and synthesized.

2.6.1 Synthesis of 4-methylphosphonophenyl Isothiocyanate (PPITC)

4-methylphosphonophenylisothiocyanate was synthesized to selectively add a phosphate chromophore at all free amines including lysine side chains and the N-terminus of digested peptides to increase their energy absorption of 10.6 μm wavelength. To synthesize PPITC, 4-Bromophenylisothiocyanate was refluxed for 30 minutes with 2x triethylphosphite after which the product was purified on a silica column with 50:50 DCM:MeOH and dried to isolate the diethyl ester form. The ester form was converted to the acidic form by reaction with 3.5 equivalents of bromotrimethylsilane which was added drop-wise and allowed to stir at room temperature for 24 hours. The reaction mixture was dried on a rotovap, and the product was dissolved in minimal MeOH. This product was allowed to stir at room temperature for an additional four hours and then dried on a rotovap and concentrated under reduced pressure.

2.6.2 Synthesis of 2, 5,-dioxo-1-((4-((2-(pyridine-2-yl)hydrazono)methyl)benzoyl) oxy) pyrrolidine-3-sulfonic acid (NN)

The surface accessibility reagent, NN, was synthesized by by refluxing 7.3 mmol of 2-hydrazinopyridine and 9.7 mmol of 4-formylbenzoic acid in 20 mL of toluene for four hours. A yellow precipitate was isolated by vacuum filtration and washed with

toluene and ethyl acetate (82.0%). One equivalent of the isolated hydrazonal intermediate was coupled with two equivalents of N-hydroxysulfosuccinimide using two equivalents of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide and a catalytic amount of 4-dimethylaminopyridine in dimethylformamide for 24 hours. Upon completion, the reaction was rotovaped and washed with dichloromethane and water. The product was titurated in hexanes before isolating a dark yellow solid (46.8%).

2.7 SELECTIVE DERIVATIZATION

2.7.1 Free Amine Modifications

Free amines are prime canidates for peptide/protein modification because they occur regularly (lysine side chains and N-terminus) and are broadly susceptible to a number of reactions.

2.7.1.1 Phenyl Isothiocyanate (PITC) and 4-Sulfophenyl Isothiocyanate (SPITC)

20 µL of a 10 µg/µL solution of reagent solution buffered to a pH of 9.5 with 20 mM $NaHCO_3$ was incubated with 10 µL of 1 mM peptide solution for 30 minutes at 55º C. The reaction was terminated with an addition of 1.5 µL of 5% trifluoroacetic acid, and then derivatized product was purified on a C18 spin column.

2.7.1.2 4-methylphosphonophenyl Isothiocyanate (PPITC)

20 µL of a 10 µg/µL solution of reagent solution buffered to a pH of 9.5 with 20 mM $NaHCO_3$ was incubated with 10 µL of 1 mM peptide solution, 60 µL 100 mM

$NaHCO_3$ (pH 8 – 9) and 70 µL water for 30 minutes at 55º C. The product was purified on a C18 spin column and diluted to 10 µM prior to ESI analysis.

2.7.1.3 2,5,-dioxo-1-((4-((2-(pyridine-2-yl)hydrazono)methyl)benzoyl)oxy) pyrrolidine-3-sulfonic acid (NN)

Each protein (1 mM, 25µL) was mixed with NN at a variety of molar ratios relative to NN (20 mM) in PBS buffer at pH 7.2 - 7.4. The protein:NN ratios were varied from 1:2 to 1:10. All reactions were carried out at room temperature overnight and cleaned up using 10 kDa MWCO filters the following morning.

2.7.2 Cysteine Alkylations

All disulfide bonds in the peptides or proteins were first reduced by adding a 5-fold molar excess of dithiothreitol (DTT) to the peptide or protein (10 µL, 1 mM) in aqueous solution and incubating at 40 °C for 1 hr in the dark (per manufacturer's recommended protocol for optimal efficiency). After alkylation all peptides were desalted using Thermo Fisher PepClean C_{18} spin columns and diluted for MS analysis as described above.

2.7.2.1 Iodoacetamide (IAM)

IAM-alkylation was performed by adding iodoacetamide (4 µL, 1 M) in ammonium bicarbonate buffer (pH ~8.0, 100 mM). This reaction was allowed to proceed at room temperature for 1 hr in the dark, followed by the addition of excess DTT to

quench the alkylation reaction. The resulting products are ones in which the reduced thiols are converted to *S*-carboxamidomethyl groups.

2.7.2.2 Dimethyl Lysine (DML)

DML modifications were initiated by reacting N,N-dimethyl-2-chloro-ethylamine (1 μL, 1 mM in 1M HEPES buffer, pH 7.8) for three hours at room temperature, producing carboxyamidomethyl cysteines.

2.7.2.3 (3-acrylamidopropyl)-trimethyl ammonium chloride (APTA)

APTA derivatization was achieved by reacting the reduced peptide solution with APTA (16 μL, 1 mg/mL in 50:50 ACN/ 100 mM ammonium bicarbonate buffer, pH 8.0) overnight at room temperature in the dark followed by freezing to terminate the reaction. The resulting products are fixed-charge trimethylammonium *S*-alkyl cysteine derivatives.

2.8 ENZYMATIC DIGESTION

2.8.1 Trypsin

Trypsin digestion produces enzymatically cleaved peptides with a C-terminal lysine or arginine amino acid when proline is not the neighboring residue. Due to the abundance of lysine and arginine residues, trypsin produces smaller peptides that upon ESI normally reside in the 1+, 2+ and 3+ charge states. Proteins diluted in 100 mM NH_4HCO_3 are digested using 1 mg/mL proteomics grade trypsin (Sigma-Aldrich, St. Louis, MO) in 1 mM HCl at a 1:100 w/w ratio overnight at 37 ^{o}C. Reactions were all

terminated with a small addition of 1.5 μL 5% trifluoroacetic acid and diluted to 10 μM before ESI analysis with 49.5/49.5/1 H_2O/MeOH/acetic Acid.

2.8.2 ArgC

ArgC digestion leads to cleavage on the C-terminal side of arginine residues. By eliminating cleavage following lysine residues ArgC produces much larger peptides to be analyzed by ESI. Proteins are diluted in 100 mM NH_4HCO_3 with a small addition of Ca acetate to promote efficient digestion and mixed with 1 mM ArgC solution at a 1:16 protein to enzyme w/w ratio. Reactions are allowed to go overnight at 37 ^{o}C. Digested samples are diluted to 10 μM before ESI analysis with 49.5/49.5/1 H_2O/MeOH/acetic acid.

2.8.3 GluC

Endoproteinase GluC is a serine protease that cleaves selectively on the C-terminal side of glutamic acid. Digestions are carried out in 50 mM Tris-HCl, 0.5 mM Glu-Glu buffer at pH 8.0 at a 1:20 protein to enzyme w/w ratio overnight at 25 ^{o}C. Digested samples are diluted to 10 μM before ESI analysis with 49.5/49.5/1 H_2O/MeOH/acetic Acid.

2.9 LIQUID CHROMATOGRAPHY

For complex samples, such as a protein digest, separation of the mixture is often advantageous prior to mass spectrometric analysis to increase the sensitivity. Most commonly reversed phase liquid chromatography is used for such samples. Liquid

chromatography separates and desalts complex samples based on the polarity and size of the different analytes and is easily adapted to mass spectral analysis. A Dionex Ultimate 3000 cap/nano HPLC system (Sunnyvale, CA) with a reverse phase Agilent ZORBAX 300SB-C_{18} column (Santa Clara, CA) (150 × 0.3 mm, 5 µm particle size) was used for all separations. The column temperature was kept constant throughout the run at 40°C. Eluent A consisted of 0.1% formic acid in water and eluent B was 0.1% formic acid in acetonitrile. A linear gradient from 5% eluent B to 40% eluent B over 65 min at 5 µL/min was used for all capillary-LC runs while 0.3 µL was used for nano-LC runs. Injections of approximately 1 µg were used for all samples.

2.10 DATABASE SEARCHING

In most proteomic applications using LC-MS/MS analysis, the vast amount of data collected is subjected to automated database searching to identify the sample. Typically, the experimental MS/MS spectra are searched against a library of theoretical spectra that are generated from known protein sequences and then scored based on how well they matched up and the ability to distinguish between true positives and false positives. SEQUEST was used for all database searching experiments through the ThermoFisher Scientific Proteome Discoverer 1.0 software. For all SEQUEST calculations the following parameters were used; signal/noise ratio of 3, a precursor mass tolerance of 1.5 Da, a fragment mass tolerance of 0.8 Da and all peptide hits were filtered against a false discovery rate (FDR) of 1%. All experimental derivatizations were set as dynamic modifications based on their associated mass shifts. Oxidation of methionines was set as a dynamic side chain modification in all experiments. For UVPD experiments,

a, *b*, *c*, *x*, *y*, and *z* product ions were used, for ETD experiments *c* and *z* product ions were used and for CID experiments *b* and *y* product ions were used.

2.11 REFERENCES

(1) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233-6237.

(2) McLuckey, S. A.; Goeringer, D. E. *J. Mass Spectrom.* **1997**, *32*, 461-474.

(3) McLuckey, S. A. *Eur. J. Mass Spectrom.*, *16*, 429-436.

(4) Wells, J. M.; McLuckey, S. A. *Methods in Enzymology* **2005**, *402*, 148-185.

(5) Laskin, J.; Futrell, J. H. *Mass Spectrometry Reviews* **2005**, *24*, 135-167.

(6) Dongre, A. R.; Somogyi, A.; Wysocki, V. H. *J. Mass Spectrom.* **1996**, *31*, 339-350.

(7) Chaurand, P.; Luetzenkirchen, F.; Spengler, B. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 91-103.

(8) Roepstorff, P. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

(9) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 1657-1665.

(10) Zhang, L.; Reilly, J. P. *Analytical Chemistry* **2009**, *81*, 7829-7838.

Chapter 3

Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization Experimental

3.1 OVERVIEW

A strategy for improving the sequencing of peptides by infrared multiphoton dissociation (IRMPD) in a linear ion trap mass spectrometer is described. We have developed an N-terminal derivatization reagent, 4-methylphosphonophenylisothiocyanate (PPITC), which allows the attachment of an IR-chromogenic phosphonite group to the N-terminus of peptides, thus enhancing their IRMPD efficiencies. After the facile derivatization process, the PPITC modified peptides require shorter irradiation times for efficient IRMPD and yield extensive series of *y* ions, including those of low *m/z* that are not detected upon traditional CID. The resulting IRMPD mass spectra afford more complete sequence coverage for both model peptides and tryptic peptides from cytochrome *c*. We compare the effectiveness of this derivatization/IRMPD approach to that of a common N-terminal sulfonation reaction that utilizes 4-sulfophenylisothiocyanate (SPITC) in conjunction with CID and IRMPD.

3.2 INTRODUCTION

In recent years, there has been tremendous effort devoted to the application of mass spectrometry to the field of proteomics[1, 2] in large part because of the success of tandem mass spectrometry for elucidation of primary sequences and modifications of peptides and proteins.[3-7] Several ion activation methods have been developed in this

context, including collision induced dissociation (CID),[8, 9] electron capture dissociation (ECD),[10] electron transfer dissociation (ETD),[11] pulsed-Q dissociation (PQD),[12] surface induced dissociation[13] and photodissociation (PD).[14-19] Each method has its own particular strengths and shortcomings, with CID being the most widely used due to its relatively well-understood underpinnings. However, the resulting CID mass spectra can be cluttered with redundant *b* and *y* ions, as well as ions created by uninformative losses of small organic molecules like water and ammonia. Hence, the interpretation of such data manually or via *de novo* algorithms remains challenging.[20, 21] ECD and ETD provide complementary *c* and *z* ions and have proven more useful for tracking post-translational modifications (PTMs); however, these activation methods generally offer lower dissociation efficiencies than CID and are highly dependent on the charge state of the selected precursor ion.[10, 11] PD methods, including both infrared multiphoton dissociation (IRMPD) and ultraviolet photodissociation (UVPD), have also shown promise because of their potential for high energy deposition and tunability.[14-19, 22-35] Since photoabsorption is not a collision-based process, it does not alter the kinetic energies of ions and thus minimizes any ion losses due to scattering.

Photodissociation also offers particular advantages for ion trap mass spectrometers. For example, photoactivation is independent of the trapping voltage, unlike CID. In CID, the rf voltage during ion activation influences the energy deposition of collisional activation as well as defines the lower m/z range. The potential for greater energy deposition in CID occurs at the expense of storage of lower m/z ions, ones which might be key diagnostic ions for sequencing peptides. PQD is another collision based activation method used as an alternative for alleviating the limited m/z storage problem

during CID in ion trap mass spectrometers. For PQD, the rf voltages applied during ion activation and subsequent ion trapping are quickly manipulated to avoid ejection of all the lower m/z fragment ions.[12] However, the dissociation efficiency for PQD, in terms of the conversion of precursor ions to fragment ions, is rather low. Despite this impressive array of ion activation methods, the quest for more efficient and more selective strategies continues to be an important goal in the development of mass spectrometric-based proteomics, especially in order to simplify spectral interpretation and improve the identification of post-translational modifications.

One approach entails derivatization of proteins or peptides to facilitate their enrichment in complex mixtures,[36-39] to alter their charge states to direct fragmentation pathways,[16] or to attach chromophores to enhance their photoabsorptivity.[14-16, 29, 32, 40, 41] For example, N-terminal sulfonation reactions have proven to simplify the MS/MS spectra of peptides by introducing a fixed negative charge at the N-terminus.[42-45] This sulfonate addition leads to neutralization of the normally singly charged *b* ions created upon dissociation of peptides, resulting in series of easily interpreted *y* ions. In particular, 4-sulfophenylisothiocyanate (SPITC) has shown great potential as a reagent for this process because it is inexpensive, stable, and extremely efficient in reacting with the N-terminus of peptides.[43, 45, 46] Upon CID the dominant product ion for the SPITC-derivatized peptides is typically the y_{n-1} ion due to the loss of the sulfanilic acid and the N-terminal amino acid which is catalyzed by the labile proton from the sulfonic group that promotes a specific Edman-type cleavage.[43] Another N-terminal sulfonation reagent, 2-sulfobenzoic acid cyclic anhydride, also introduces a negative charge at the N-termini of peptides, but the resulting peptides dissociate to give a broader array of *y* ions than

afforded by the SPITC-derivatized peptides.[16] The newest N-terminal sulfonation reagent, 4-chlorosulfophenyl isocyanate, also results in more uniform arrays of *y* ions from the peptides, albeit with higher critical energies.[46] PITC is another N-terminal reagent that affords a good chromophore at 266 nm for UVPD, and the resulting PITC-peptides yield both *b* and *y* sequence ions.[47] The same SPITC-derivatized peptides produce only *y* ions upon UVPD, again with notable enhancement of the y_{n-1} ion.[47]

We have been developing derivatization agents for biological molecules that increase their photoabsorptivities in the gas phase and/or alter the fragmentation patterns of the resulting ions to give more informative or more easily interpreted mass spectra. For example, we recently combined an N-terminal sulfonation procedure with IRMPD in a quadrupole ion trap to allow detection of entire series of *y* ions for tryptic peptides, thus allowing successful *de novo* sequencing.[16] We have also developed other IR-chromogenic ligands to improve IRMPD efficiencies, including phosphorylated pyridyl ligands that form metal complexes with flavonoids,[41] a boronic acid reagent that selectively binds to the diol functionality of oligosaccharides and improves the determination of their sequences,[40] and a new phosphorylated crosslinking reagent that allows the differentiation of crosslinked peptides from non-crosslinked peptides for elucidation of contact points in proteins.[48] We have also explored UV-chromogenic reagents in conjunction with UVPD for sequencing peptides[15] and a reductive amination procedure that incorporates fluorescent labels into oligosaccharides to improve UVPD efficiencies.[14] The successful outcomes of these previous strategies that combined derivatization with photodissociation have motivated our efforts to expand the arsenal of targeted chromogenic reagents for tunable photodissociation of biological molecules.

In this chapter we report a new N-terminal derivatization reagent that incorporates a phosphonite functionality. The new reagent, 4-methylphosphonophenylisothiocyanate (PPITC), was designed specifically to provide a negative charge site similar to the sulfonated reagents, in addition to containing a strong IR chromophore to enhance the photoabsorptivity at a wavelength of 10.6 μm of the resulting derivatized peptides. It has been shown previously that phosphorylated peptides possess high photoabsorptivities at 10.6 μm which is critical for successful IRMPD.[23, 49-51] Here we compare the IRMPD efficiencies and the number of diagnostic sequence ions obtained for a series of peptides derivatized by SPITC, PPITC, and PITC (Figure 1).

3.3 EXPERIMENTAL

3.3.1 Materials and Reagents

All chemicals were purchased from Sigma Aldrich (St. Louis, MO) except the following peptides which were obtained from BACHEM (King of Prussia, PA): LQVQLSIR (H-4588), FSWGAEGQR (H-2680), Amyloid protein 1-16, and YGGFLK (H-2765) and from American Peptide Co. (Sunnyvale, CA): HSDAVFTDNYTR (48-1-12) and Fibrinopeptide A (42-1-12A).

3.3.2 Synthesis of 4-methylphosphonophenylisothiocyanate (PPITC)

4-Bromophenylisothiocyanate was refluxed for 30 minutes with 2x triethylphosphite after which the product was purified on a silica column with 50:50 DCM:MeOH and dried to isolate the diethyl ester form. The ester form was converted to the acidic form by reaction with 3.5 equivalents of bromotrimethylsilane which was

added drop-wise and allowed to stir at room temperature for 24 hours. The reaction mixture was dried on a rotovap, and the product was dissolved in minimal MeOH. This product was allowed to stir at room temperature for an additional four hours and then dried on a rotovap and concentrated under reduced pressure. The PPITC structure was confirmed by ^{1}H NMR using a 400 MHz Varian instrument. ^{1}H NMR at 400 MHz in CD_3OD yielded: 7.21 ppm (d, 2H), 7.40 ppm (d, 2H), 3.17 ppm (s, 2H), 3.19 ppm (s, 1H); these findings are consistent with the proposed structure.

3.3.3 PITC Derivatization of Peptides

PITC and SPITC protocol: 20 µL of a 10 µg/µL solution of reagent solution buffered to a pH of 9.5 with 20 mM $NaHCO_3$ was incubated with 10 µL of 1 mM peptide solution for 30 minutes at 55º C. The reaction was terminated with an addition of 1.5 µL of 5% trifluoroacetic acid, and then derivatized product was purified on a C18 spin column.

PPITC protocol: 20 µL of a 10 µg/µL solution of reagent solution buffered to a pH of 9.5 with 20 mM $NaHCO_3$ was incubated with 10 µL of 1 mM peptide solution, 60 µL 100 mM $NaHCO_3$ (pH 8 – 9) and 70 µL water for 30 minutes at 55º C. The product was purified on a C18 spin column and diluted to 10 µM prior to ESI analysis.

3.3.4 Mass Spectrometry

All experiments were performed on a ThermoFinnigan LTQ XL linear ion trap mass spectrometer equipped with a standard ESI source. The LTQ mass spectrometer was modified for IRMPD as described earlier.[48] Solutions were directly infused at a flow

rate of 3 µL/min at a concentration of 10 µM in 100% MeOH. For CID, the precursor was activated for 30 ms at a standard q_z value of 0.25. Energy-variable CID experiments were performed by incrementally raising the applied CID voltage. All CID experiments were recorded in triplicate on the same day. In cases where CID voltages are reported, the CID voltages required to cause dissociation of 50% of the selected precursor ions were corrected for their number of degrees of freedom using the following equation:

Equation 3.1:
$$E_{(50\%corrected)} = E_{(50\%)} \times \frac{N_{u\,\mathrm{mod}\,ified}}{N_{derivatized}}$$

where the degrees of freedom is defined as N = 3*n* – 6, and *n* is the number of atoms in the underivatized peptide.[52]

3.3.5 Infrared Multiphoton Dissociation

IRMPD experiments were performed using a 48-5 Synrad 50-W continuous wave CO2 laser (Mukilteo, WA). The laser was directed through the back end of the vacuum manifold that was modified with a CF view port flange equipped with a ZnSe window to allow the 10.6 µm radiation into the ion trap.[48] This port was aligned so that the beam was on axis with a 1.7 mm entrance aperture and the 2 mm aperture of the exit lens of the ion trap. The unfocused laser was controlled using a TTL signal setup in the instrument software to trigger during activation. All IRMPD experiments were performed using a reduced q_z value of 0.1 to allow for sufficient detection of the low-mass ions and irradiation times of ~5 to 200 ms at 50 W unless otherwise noted. The helium pressure was maintained at the normal operating pressure. Time-variable IRMPD experiments were executed by incrementally increasing the irradiation time.

3.4 RESULTS AND DISCUSSION

3.4.1 Design and Reactions of PPITC

We have previously shown that combining IRMPD with an N-terminal sulfonation procedure substantially improved the *de novo* sequencing of peptides, primarily by eliminating the redundant *b* ion series that can make it difficult to pinpoint the *y* ions and also by extending the series of *y* ions detected via alleviation of the low mass cut-off typically encountered upon conventional CID.[16] The promising results obtained in this prior study motivated us to design a new N-terminal derivatization reagent that would specifically contain an IR chromophore to enhance the absorption of 10.6 μm photons, thus making the peptides even more amenable to IRMPD. In addition, the new reagent was designed to afford a fixed negative charge site at the N-terminus to encourage the neutralization of all *b* fragments, yielding *y*-rich fragmentation patterns similar to what has been seen previously following derivatizations of peptides with SPITC.[43, 45, 46] The negative phosphonite group fixed at the N-terminus would neutralize any N-terminal ions created through tandem excitation leaving solely C-terminal ions represented in the resulting MS/MS spectrum, which for both CID and IRMPD is predominantly the y ion cleavage. The phosphono-PITC reagent shown in **Figure 3.1** resulted in the attainment of both goals. The phosphonite group was incorporated on the PITC skeleton via a two-step modification of 4-bromophenylisothiocyanate. Subsequent reaction of PPITC with individual peptides or mixtures of tryptic peptides occurred with yields of 40-90% percent, thus resulting in formation of ample PPITC-derivatized

peptides for IRMPD. In comparison, the average yields for derivatization of the same peptides using PITC or SPITC were typically 20-40% and 70-90%, respectively.

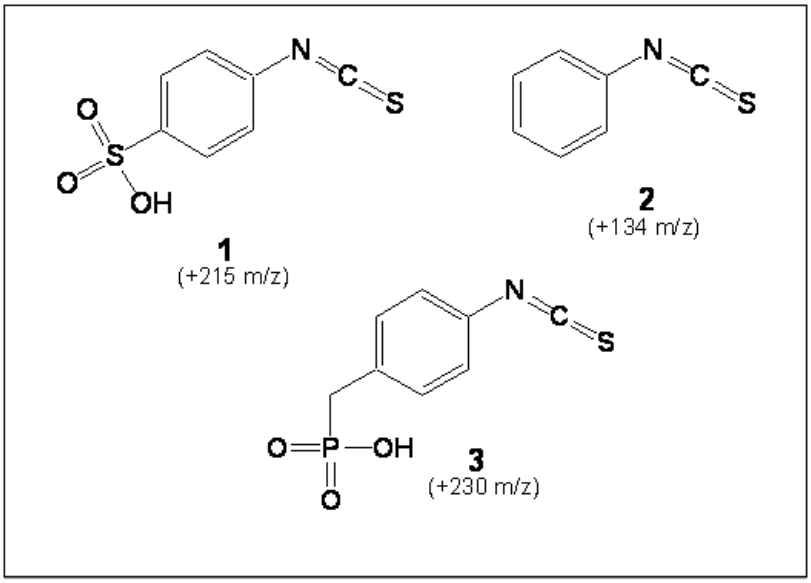

Figure 3.1 PITC complexes; (**1**) 4-Sulfophenylisothiocyanate (SPITC), (**2**) Phenylisothiocyanate (PITC), (**3**) 4-Methylphosphonophenylisothiocyanate (PPITC)

3.4.2 CID and IRMPD of PITC-Derivatized Peptides

Examples of the CID and IRMPD mass spectra obtained for the underivatized and the corresponding PPITC-derivatized peptides are shown in **Figure 3.2** and **Figure 3.3**. The CID mass spectrum of the first protonated peptide, FSWGAEGQR, is dominated by two fragment ions due to dehydration, and the diagnostic sequence ions have low relative

abundances (**Figure 3.2A**). Moreover, the spectrum is cluttered with both *b* and *y* ions that offer redundant and yet incomplete sequence information, in addition to additional confounding fragment ions attributed to ammonia or water losses. IRMPD of the underivatized peptide is very inefficient, resulting in no dissociation (spectrum not shown). Following the derivatization of this peptide with the PPITC reagent, IRMPD results exclusively in a complete series of *y* ions from y_1 to y_8 (**Figure 3.2C**). CID of the PPITC-peptide (**Figure 3.2B**) results in the formation of the y_{n-1} as the most abundant sequence ion (i.e. y_8), a finding similar to what has been previously reported for CID of SPITC-derivatized peptides.[46, 53, 54] CID of the modified peptide also results in the formation of the $[M + CS + H]^+$ which represents the partial loss of the PPITC reagent (-188 Da). This could serve as a derivatization marker since all modified peptides should yield a fragment resulting from the loss of 188 Da from the precursor ion and could provide a simple test to distinguish between derivatized and underivatized peptides.

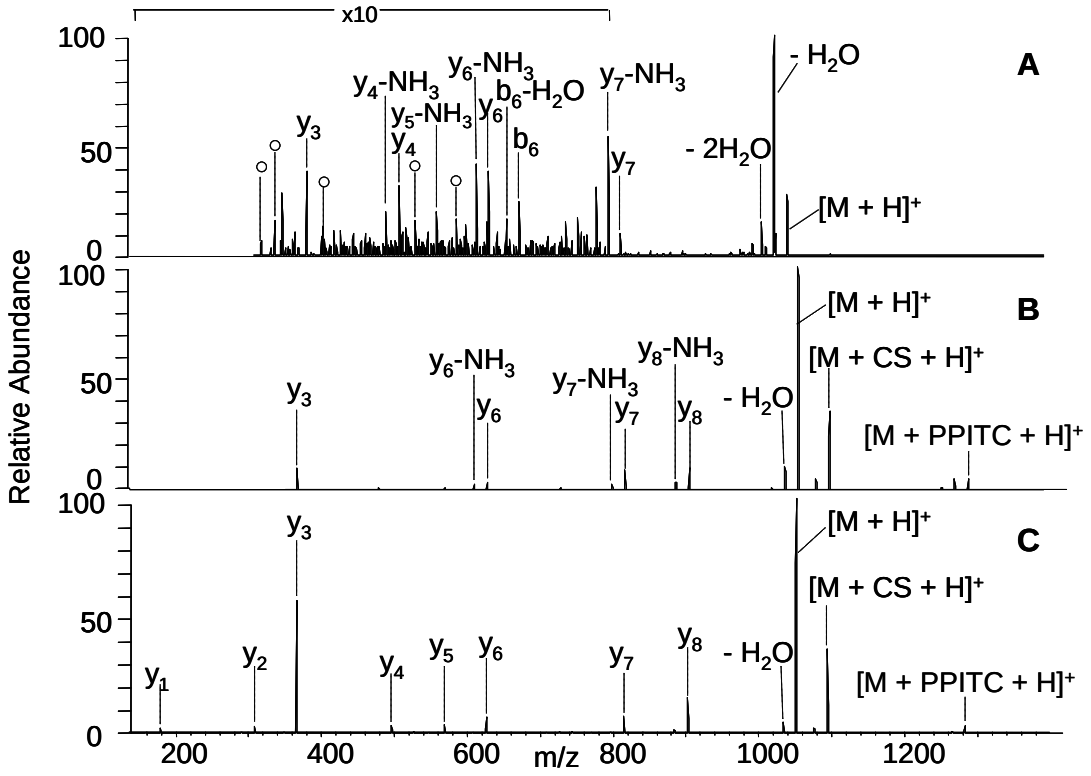

Figure 3.2 ESI MS/MS mass spectra of protonated FSWGAEGQR, (**A**) CID of unmodified peptide (CID voltage 0.071 V, 30 msec), (**B**) CID of PPITC-modified peptide (CID voltage 0.087V, 30 msec), (**C**) IRMPD of PPITC-modified peptide (laser power 50 W, 20 msec). The $[M + CS + H]^+$ ion represents the partial loss of PPITC reagent. Magnification applies to all spectra over the mass range indicated. Water losses are represented by the pound sign and internal products by ○.

The CID and IRMPD mass spectra obtained for a larger peptide, fibrinopeptide A (ADSGEGDFLAEGGGVR), are shown in **Figure 3.3**. The CID mass spectrum of the unmodified protonated peptide (**Figure 3.3A**) is dominated by two *y* ions (y_9 and y_{14}) in addition to a few other *y* ions of low abundance, dehydrated *y* ions, and internal fragment ions of low abundance. This spectrum is unsatisfactory for *de novo* sequencing or manual interpretation, especially if the primary protein sequence of origin is unknown prior to

analysis. IRMPD of the unmodified protonated peptide again results in no dissociation (spectrum not shown), indicating that the peptide exhibits poor IR absorptivity and undergoes collisional deactivation more efficiently than energization. CID of the PPITC-derivatized peptide results in the creation of different y ions than detected for the underivatized peptide (y_9, y_{11}, y_{13}, y_{14}, y_{15}) but leaves the sequence of the entire C-terminal side of the peptide inconclusive (**Figure 3.3B**). Moreover, gaps in the second half of the *y* series could lead to ambiguities in identifying the amino acid sequence or the incorrect assignment of residues. On the other hand, the IRMPD mass spectrum of the PPITC-derivatized peptide displays an extensive series of *y* ions ranging from y_2 to y_{15} (**Figure 3.3C**), making sequence interpretation for this peptide straightforward. The high absorptivity of the derivatized peptide is reflected by the short irradiation time (10 msec) needed to obtain the IRMPD spectrum shown in **Figure 3.3C**. These short irradiation times would also provide a simple, quick test for determining derivatized peptides from unmodified peptides because only derivatized peptides will show any fragmentation with such a short irradiation time.

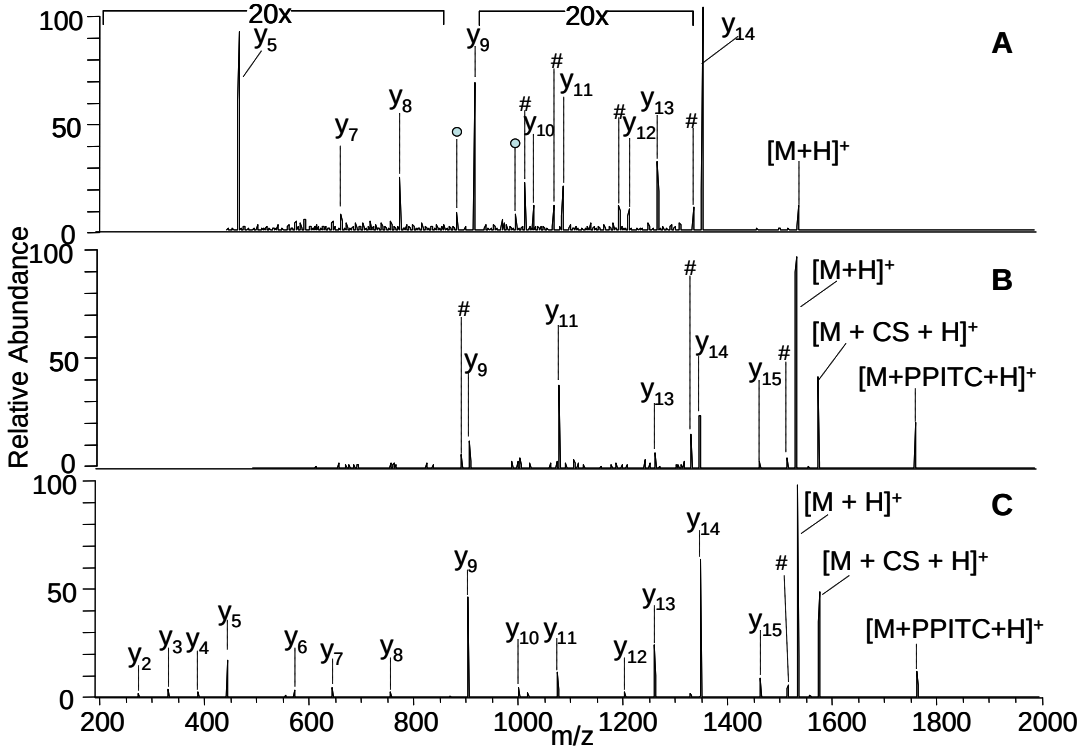

Figure 3.3 ESI MS/MS of protonated Fibrinopeptide A; (**A**) CID of unmodified peptide (CID voltage 0.11 V, 30 msec), (**B**) CID of PPITC-modified peptide (CID voltage 0.13 V, 30 msec), (**C**) IRMPD of PPITC-modified peptide (laser power 50 W, 10 msec). The $[M + CS + H]^+$ ion represents the partial loss of PPITC reagent. Water losses are represented by # and internal fragments by ○. Magnification applies to all spectra over mass range indicated.

As shown by Muddiman et al.,[49-51] phosphorylation greatly increases IR absorptivities of peptides and this finding is consistent with our results for the PPITC-derivatized peptides. To further probe the relative absorptivities of PITC-derivatized peptides, we examined the IRMPD mass spectra of peptides modified by PITC, SPITC, and PPITC. Representative results are shown in **Figure 3.3** for one peptide, FSWGAEGQR. The PITC-derivatized peptide exhibits little dissociation even upon 500

msec of irradiation at 10.6 μm (**Figure 3.4A**), similar to what was observed for non-derivatized peptides and giving a clear indication of the poor absorptivity of the PITC-peptides. IRMPD of the SPITC-peptide results in the dominant production of the y_{n-1} ion (in this case y_8) after 25 msec of irradiation (**Figure 3.4B**). In contrast, the entire series of *y* ions is observed in the IRMPD mass spectrum of the corresponding PPITC-peptide after 20 msec (**Figure 3.4C**). Both the SPITC- and PPITC-derivatized peptides appear to have comparable absorptivities, but the PPITC-peptide yields a more complete series of *y* ions.

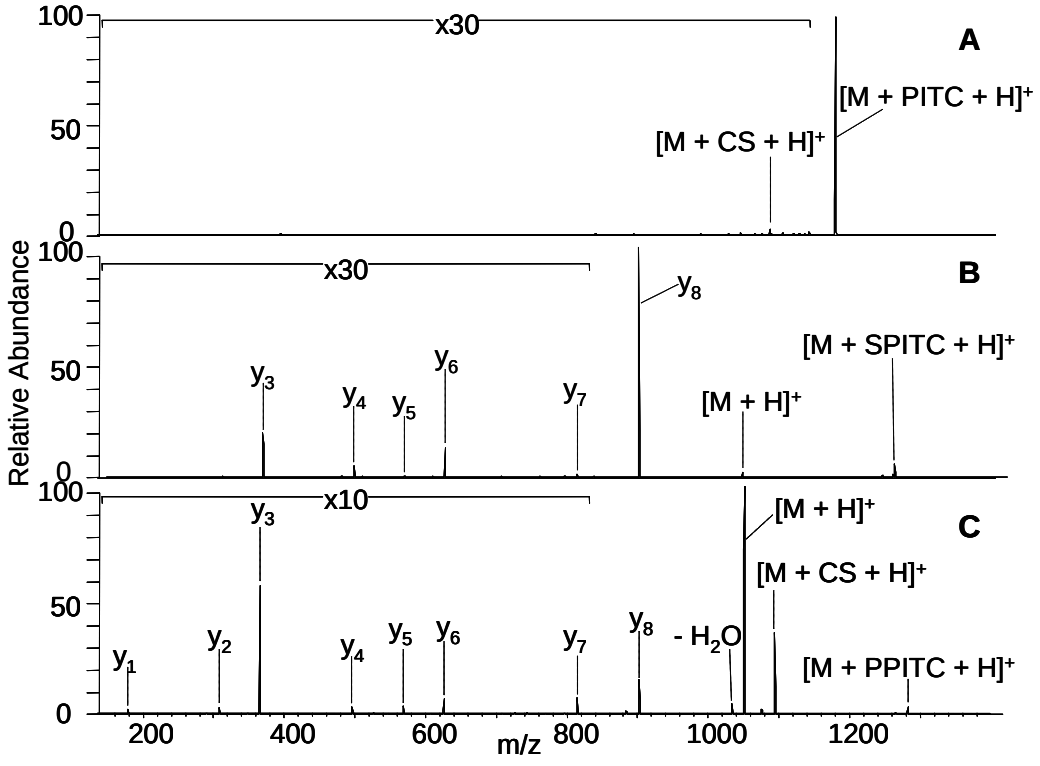

Figure 3.4 IRMPD of protonated derivatized FSWGAEGQR; (**A**) PITC-peptide (laser power 50 W, 500 msec), (**B**) SPITC-modification (laser power 50 W, 25 msec), (**C**) PPITC modification (laser power 50 W, 20 msec). The $[M + CS + H]^+$ ion represents the partial loss of derivatization reagent.

CID and IRMPD mass spectra were acquired in triplicate for a series of eight PITC-, SPITC-, PPITC-derivatized peptides, as well as the underivatized peptides. The results are summarized in bar graph form in **Figure 3.5** in terms of sequence coverage percentage, which is defined as the number of *y* sequence ions detected relative to the total number of possible *y* sequence ions (y_1 to y_{n-1}), expressed as a percentage. For this comparison, a *y* ion is considered to be detected in a CID or IRMPD mass spectrum if it has a total peak area of 1% or greater. A value of 100% for the sequence coverage percentage means that every possible y ion (y_1 to y_{n-1}) is detected.

In most cases, the highest sequence coverage percentages were associated with the PPITC-derivatized peptides in conjunction with IRMPD, suggesting that IRMPD of the PPITC-peptides affords the optimal strategy for identification of peptide sequences. The main exception to this trend occurs for multiply charged ions (i.e. doubly protonated VIP and multiply protonated amyloid peptide (DAEFRHDSGYEVHHEK) (data not shown)). For the multiply protonated peptides, the extra mobile proton allows both a/*b* and *y* ions to remain charged in the MS/MS spectra and thus results in both a more elaborate array of potential a/*b* and *y* ions and overall lower abundance *y* ions amongst a more cluttered spectra. For YGGFLK, the sequence coverage percentage obtained upon CID of the unmodified peptide is comparable to that obtained upon IRMPD of the PPITC-derivatized peptide. However, the CID mass spectrum contains both *b* and *y* ions which complicates the spectral interpretation, whereas the IRMPD spectrum displays only *y* ions.

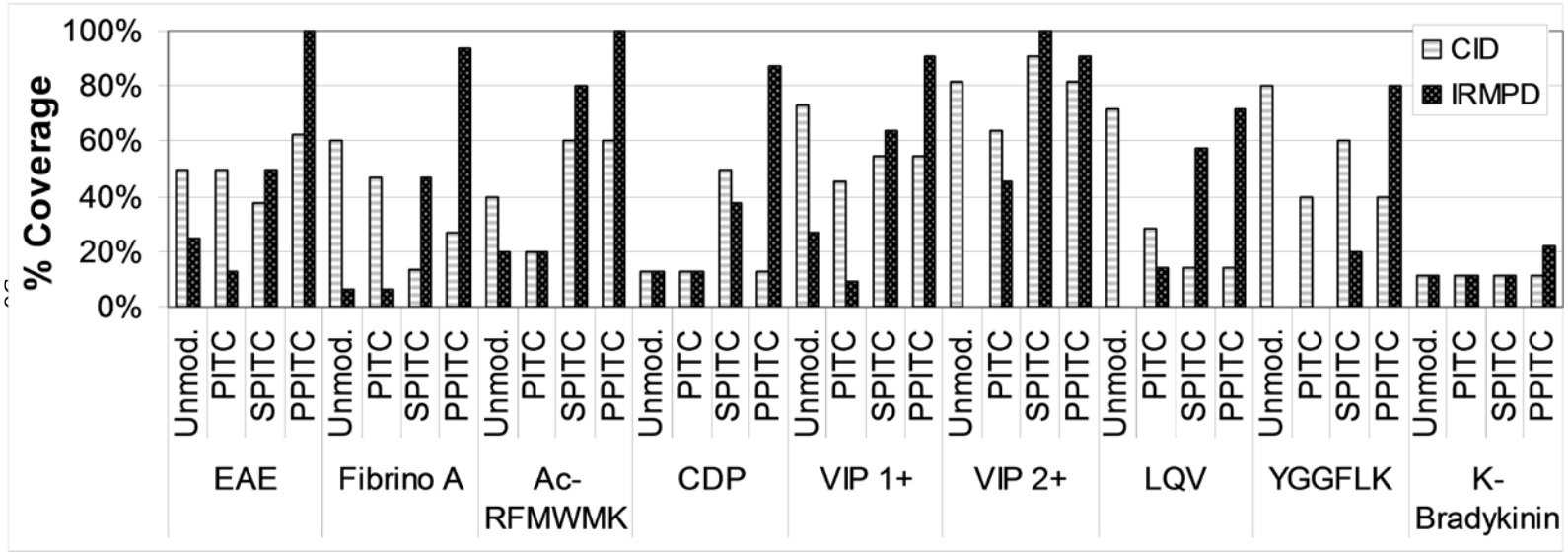

Figure 3.5 Sequence coverage percentages obtained by CID and IRMPD of unmodified and modified peptides, where the percent coverage is defined as the number of *y* sequence ions detected relative to the total number of possible *y* sequence ions (y_1 to y_{n-1}), expressed as a percentage.
Unmod.= unmodified peptide, EAE = FSWGAEGQR, Fibrino A = ADSGEGDFLAEGGGVR, CDP = CDPGYIGSR, VIP = HSDAVFTDNYTR, LQV = LQVQLSIR, K-Brady = KRPPGFSPFR.
*For AcRFMWMK, the percent coverage is defined as the number of *b* sequence ions detected relative to the total number of possible *b* sequence ions.

3.4.3 Energy-Variable CID and IRMPD

We previously found that the improvement in IRMPD efficiency afforded by the N-terminal sulfonation of peptides (using 2-sulfobenzoic acid) was due in part to a significant decrease in the critical energies of the peptides.[16] For the present study, the CID voltages and IRMPD irradiation times needed to cause dissociation of 50% of the parent ion population were determined and are summarized in **Table 3.1** for one representative peptide, fibrinopeptide A. The required CID voltages do not change significantly for the unmodified, PITC-, SPITC-, and PPITC-derivatized peptides, suggesting that there is not a substantial change in the critical energies. In contrast, there is a notable change in the IRMPD irradiation times for the same series of peptides, with the SPITC- and PPITC-derivatized peptides requiring less than 25 msec whereas the underivatized and PITC-derivatized peptides never reaching the 50% dissociation level even after 1000 msec of irradiation time. The results suggest that the SPITC- and PPITC-derivatized peptides possess significantly amplified IR absorptivities which make them more amenable to IRMPD.

Peptide	IRMPD irradiation time (ms)	CID voltage (V)
Unmodified	> 1000	0.104
PITC	> 1000	0.107
SPITC	11.5	0.109
PPITC	9.0	0.106

Table 3.1 IRMPD irradiation times and CID voltages required to cause 50% dissociation of unmodified and modified-Fibrinopeptide A (protonated). All IRMPD irradiation values are ±1 ms and all CID values are ± 0.001 V based upon triplicate measurements.

It is also informative to consider the distributions of fragment ions obtained for the PPITC-derivatized peptides as a function of the CID voltage or IRMPD irradiation time. Monitoring the fragment ion distributions allows the genealogies of ions to be indirectly evaluated. Maps of the fragment ion distributions are shown in **Figure 3.6** for one representative peptide, PPITC-derivatized fibrinopeptide A. The CID data (**Figure 3.6A**) shows that all of the *y* ions evolve simultaneously, and their relative contributions do not change once the population of the protonated peptide is for the most part annihilated. The relative abundances of the various *y* ions continue to change as the irradiation time is increased for the IRMPD data (**Figure 3.6B**), even after the abundance of the parent ion is depleted. For example, the abundances of the y_4 and y_9 ions increase while the abundances of the y_{14} and y_{15} ions decrease at longer irradiation times. These complementary shifts in ion abundances support genealogical relationships between these sets of *y* ions, and this reflects the non-resonant nature of IRMPD that promotes conversion of primary fragment ions to secondary fragment ions, thus leading to the greater array of diagnostic *y* sequence ions.

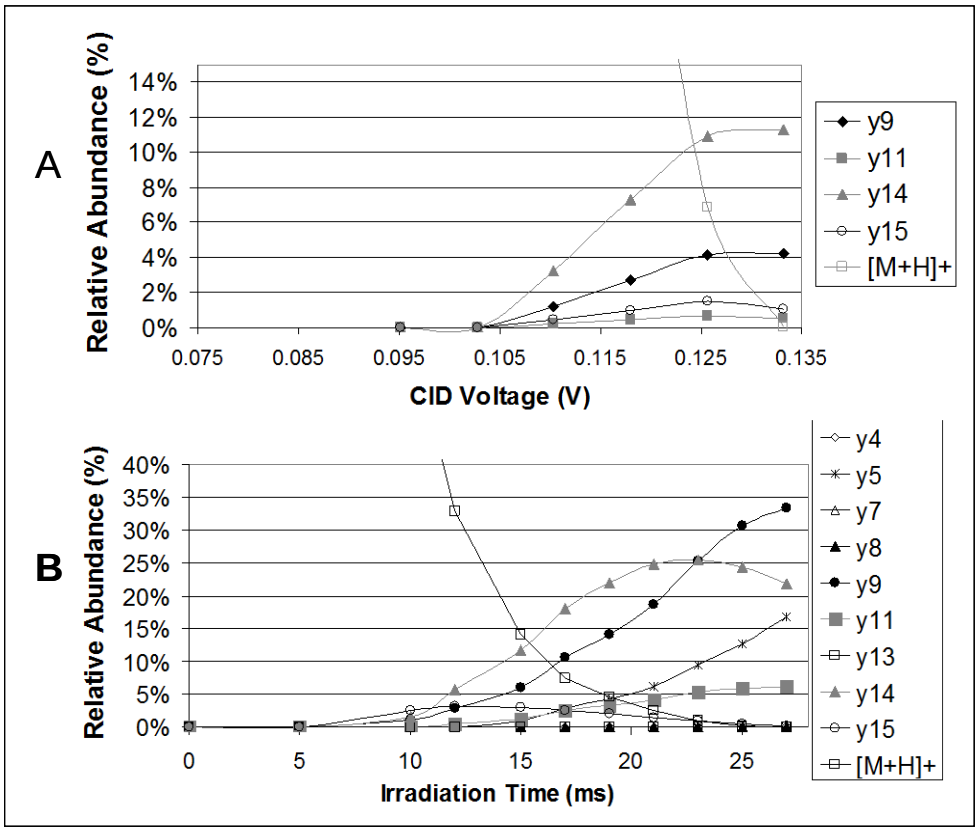

Figure 3.6 Energy-variable MS/MS distributions for y ions of protonated PPITC-Fibrinopeptide A and the surviving protonated precursor: (**A**) energy-variable CID and (**B**) time-variable IRMPD

3.4.4 PPITC Derivatization and IRMPD of Tryptic Peptides

The feasibility of the PPITC/IRMPD strategy was demonstrated for analysis of peptides obtained from the tryptic digestion of a protein, cytochrome c. The crude mixture of peptides obtained from enzymatic digestion of cytochrome c was derivatized with PPITC as described earlier, and then subjected to C18 clean-up prior to ESI-MS/MS analysis via direct infusion. A typical ESI mass spectrum of the PPITC-derivatized tryptic digest is shown as **Figure 3.7**. The majority of the peaks are assigned readily as

underivatized peptides and their corresponding PPITC-derivatized analogs based on the characteristic mass shift associated with the PPITC modification. IRMPD, with an average irradiation time of 50 msec at 50 W, and CID, with an average voltage of 0.58 V for 30 msec, were used to analyze the PPITC-derivatized peptides, and the sequence coverage percentages are summarized in **Figure 3.8**. The sequence coverage percentages obtained by IRMPD are consistently higher or at least comparable to the values obtained by CID, thus demonstrating the feasibility of applying the PPITC/IRMPD approach for proteomics. For example, upon CID the PPITC-derivatized tryptic peptide MIFAGIK produced the y_3, y_4, and y_5 ions whereas the y_2, y_3, y_4, y_5, and y_6 ions were obtained upon IRMPD. For the sole tryptic peptide with a C-terminal arginine, an even greater increase in coverage was obtained by IRMPD. Following CID, only the y_5 and y_6 ions were observed for the 1^+ charge state while the y_2, y_3, and the y_5-y_{10} were observed for the 2^+ charge state. In comparison, IRMPD of the 1^+ peptide yielded the y_2 through y_7 ions, as well as the y_9 and y_{10} ions, and IRMPD of the 2^+ peptide generated the y_2, y_3, y_5, y_6, y_9, and y_{10} ions.

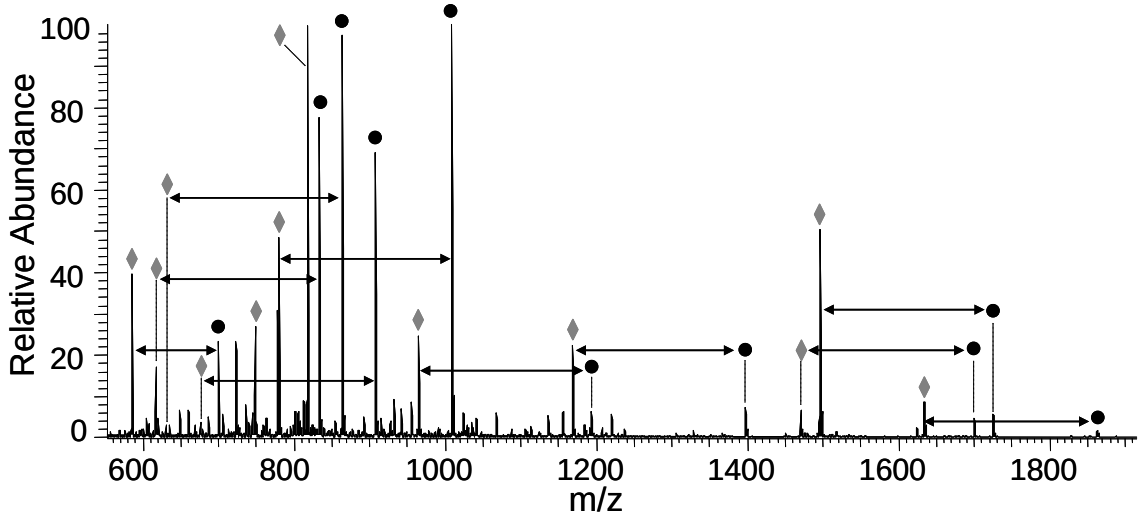

Figure 3.7 PPITC derivatization of the tryptic digest of cytochrome C. ● = PPITC derivatized tryptic peptides, ◊ = unmodified tryptic peptides. The arrows link the unmodified peptides with their PPTIC-derivatized counterparts.

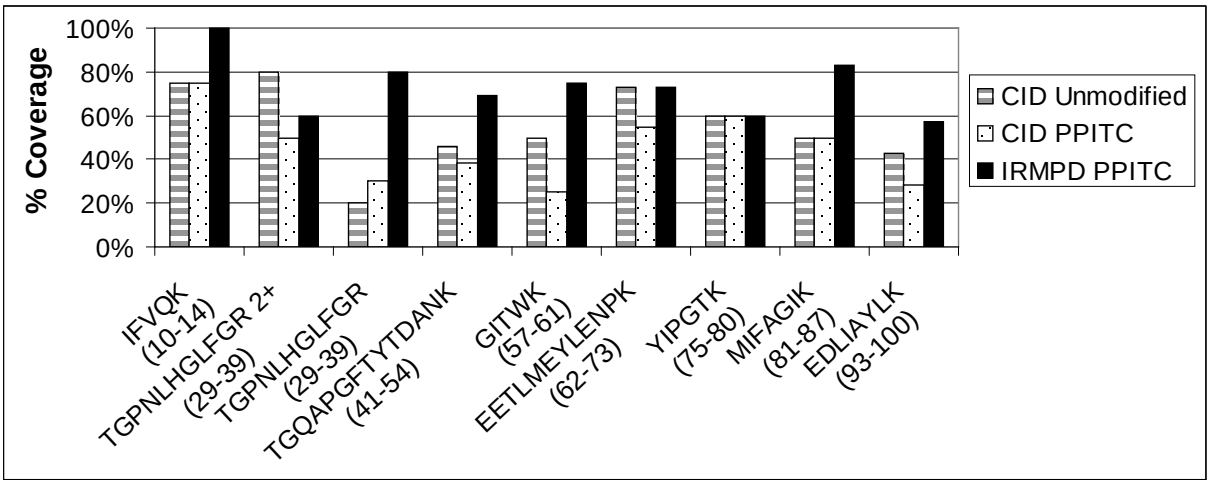

Figure 3.8 Sequence coverage percentages obtained by CID and IRMPD of unmodified and PPITC-modified tryptic peptides from cytochrome c, where the percent coverage is defined as the number of *y* sequence ions detected relative to the total number of possible *y* sequence ions (y_1 to y_{n-1}), expressed as a percentage.

In general, the overall improvement in the sequence coverage percentages obtained upon IRMPD in comparison to CID of the PPTIC-derivatized tryptic peptides is not as great as was seen for the individual model peptides summarized in **Figure 3.5**, and this is attributed to several reasons. First, the relative complexity of the tryptic digest mixture leads to lower yields of PPITC-modified tryptic peptides and corresponding lower CID and IRMPD sensitivities. Second, the formation of predominantly lysine-terminated tryptic peptides in some cases leads to competitive formation of *a/b* ions that divides the ion current among *a*, *b*, and *y* ions and ultimately reduces the sensitivity of *y* ion detection. This result may be attributed to the lower gas-phase basicity of lysine in comparison to arginine.[55, 56] This difference in basicities of the terminal residues may change the proton mobility, thus altering the possibility of creating charged *a/b* ions in comparison with the *y* ions that stem from the sequestered N-terminal proton. The sole arginine-terminated peptide, TGPNLHGLFGR, in the 1^+ charge state shows the greatest improvement in sequence coverage percentage upon IRMPD, thus supporting that the increased basicity of the arginine terminus assists in sequestering the proton on the C-terminus and thus eliminating the formation of charged *a/b* ions. However, despite the decreased basicity of lysine-terminated tryptic peptides, IRMPD analysis following PPITC modification still afforded the highest sequence coverage when compared to the other strategies (CID or using other derivatization reactions).

3.5 CONCLUSIONS

N-terminal derivatization using a novel phosphorylated phenylisothiocyanate reagent in combination with IRMPD has been shown to be a promising approach for *de*

novo sequencing of peptides. 4-Methylphosphonophenylisothiocyanate allows selective attachment of an IR chromophore to the peptides, thus enhancing their IR absorptivities and significantly reducing the irradiation times needed to cause efficient photodissociation. Extensive series of *y* ions are created upon IRMPD of the PPITC-derivatized peptides, and the use of IRMPD allows detection of a greater array of the lower m/z *y* ions than possible with conventional CID. The PPITC/IRMPD strategy gives high sequence coverage percentages and has proven viable for analysis of tryptic peptides.

3.5 REFERENCES

(1) Winston, R. J.; Fitzgerald, M. C. *Mass Spectrometry Reviews* **1997**, 16, 165-179.

(2) Godovac-Zimmerman, J.; Brown, L. R. *Mass Spectrometry Reviews* **2001**, 20, 1-57.

(3) Faux, M. C.; Scott, J. D. *Trends in Biochemical Science* **1996**, 21, 312-315.

(4) Lill, J. *Mass Spectrometry Reviews* **2003**, 22, 182-194.

(5) Regnier, F. E.; Riggs, L.; Zhang, R.; Xiong, L.; Liu, P.; Chakraborty, A.; Seeley, E.; Sioma, C.; Thompson, R. A. *Journal of Mass Spectrometry* **2002**, 37, 133-145.

(6) Sun, H.; Tonks, N. K. Trends Biochem. Sci. **1994**, 19, 480-485.

(7) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. *Methods* **2005**, 35, 211-222.

(8) Koy, C.; Mikkat, S.; Raptakis, E.; Sutton, C.; Resch, M.; Tanaka, K.; Glocker, M. O. *Proteomics* **2003**, 3, 851-858.

(9) Moyer, S. C.; Cotter, R. J.; Woods, A. S. *Journal of the American Society of Mass Spectrometry* **2002**, 13, 274-283.

(10) Zubarev, R. A. *Current Opinion in Biotechnology* **2004**, 15, 12-16.

(11) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences in the United States of America* **2004**, 101, 9528-9533.

(12) Schwartz, J. C. 2005, 2004941 653 6949743; 20040914.

(13) Dongre, A. R.; Somogyi, A.; Wysocki, V. H. *Journal of Mass Spectrometry* **1996**, 31, 339-350.

(14) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2008**, 80, 5186-5196.

(15) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2007**, 79, 7883-7892.

(16) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2006**, 78, 6855-6862.

(17) Payne, A. H.; Glish, G. L. *Analytical Chemistry* **2001**, 73, 3542-3548.

(18) Little, D. P.; Speir, J. P.; Menko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Analytical Chemistry* **1994**, 66, 2809-2815.

(19) Crowe, M. C.; Brodbelt, J. S. *Journal of the American Society of Mass Spectrometry* **2004**, 15, 1581-1592.

(20) Eng, J. K.; McCormack, A. L.; Yates, J. R. III. *Journal of the American Society of Mass Spectrometry* **1994**, 5, 976-989.

(21) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, 20, 3551-3567.

(22) Colorado, A.; Shen, J. X.; Brodbelt, J. S. *Analytical Chemistry* **1996**, 68, 4033-4043.

(23) Crowe, M. C.; Brodbelt, J. S. *Analytical Chemistry* **2005**, 77, 5726-5734.

(24) Crowe, M. C.; Brodbelt, J. S. Goolsby, B. J.; Hergenrother, P. *Journal of the American Society of Mass Spectrometry* **2002**, 13, 630-649.

(25) Goolsby, B. J.; Brodbelt, J. S. *Analytical Chemistry* **2001**, 73, 1270-1276.

(26) Hashimoto, Y.; Hasegawa, H.; Yoshinari, K.; Waki, I. *Analytical Chemistry* **2003**, 75, 420-425.

(27) Keller, K. M.; Brodbelt, J. S. *Analytical Biochemistry* **2004**, 326, 200-210.

(28) Shen, J.; Brodbelt, J. S. *Analyst* **2000**, 125, 641-650.

(29) Wilson, J. J.; Kirkovits, G. J.; Brodbelt, J. S. *Journal of the American Society of Mass Spectrometry* **2008**, 19, 257-260.

(30) Devakumar, A.; Mechref, Y.; Kang, P.; Novotny, M. V.; Reilly, J. P. *Journal of the American Society of Mass Spectometry* **2008**, 19, 1027-1040.

(31) Thompson, M. S.; Cui, W.; Reilly, J. P. *Angewandte Chemie, International Edition* **2004**, 43, 4791-4794.

(32) Ly, T.; Julian, R. R. *Journal of the American Chemical Society* **2008**, 130, 351-358.

(33) Yoon, S. H.; Chung, Y. J.; Kim, M. S. *Journal of the American Society of Mass Spectrometry* **2008**, 19, 645-655.

(34) Moon, J. H.; Shin, Y. S.; Cha, H. J. *Rapid Communications in Mass Spectrometry* **2007**, 18, 1729-1739.

(35) Yoon, S. H.; Kim, M. S. *Journal of the American Society of Mass Spectrometry* **2007**, 18, 1729-1739.

(36) Krijgsveld, J.; Heck, A. J. R. *Drug Discovery Today: Targets* **2004**, 3, S11-S15.

(37) Qu, J.; Straubinger, R .M. *Rapid Communications in Mass Spectrometry* **2005**, 19, 2857-2864.

(38) Yomota, C.; Ohnishi, Y. *Journal of Chromatography A* **2007**, 1142, 231-235.

(39) Li, H.; Schopfer, L.; Spaulding, R..; Thompson, C. M.; Lockridge, O. *Chemico-Biological Interactions* **2005**, 157/158, 383-384.

(40) Pikulski, M.; Hargrove, A.; Shabbir, S. H.; Anslyn, E. V.; Brodbelt, J. S. *Journal of the American Society of Mass Spectrometry* **2007**, 18, 2094-2106.

(41) Pikulski, M.; Wilson, J. J.; Aguilar, A.; Brodbelt, J. S. *Analytical Chemistry* **2006**, 78, 8512-8517.

(42) Keough, T.; Youngquist, R. S.; Lacy, M. P. *Proceedings of the National Academy of Sciences in the United States of America* **1999**, 96, 7131-7136.

(43) Lee, Y. H.; Shin, J.-W.; Ryu, S.; Lee, S.-W.; lee, C. H., Lee, K. *Analytical Chemistry* **2006**, 556, 140-144.

(44) Leon, I. R.; Neves-Ferreira, A. G. C.; Valente, R. H.; Mota, E. M.; Lenzi, H. L.; Perales, J. *Journal of Mass Spectrometry* **2007**, 42, 1363-1374.

(45) Wang, D.; Kalb, S. R.; Cotter, R. J. *Rapid Communications in Mass Spectrometry* **2004**, 18, 96-102.

(46) Shin, J.-W.; Lee, Y. H.; Hwang, S.; Lee, S.-W. *Journal of Mass Spectrometry* **2007**, 42, 380-388.

(47) Oh, J. Y.; Moon, J. H.: lee, Y. H.; Hyung, S.-W.; Lee, S.-W.; Kim, M. S. *Rapid Communications in Mass Spectrometry* **2005**, 19, 1283-1288.

(48) Garner, M. W.; Vasicek, L. A.; Shabbir, S.; Ansyln, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, 80, 4807-4819.

(49) Flora, J. W.; Muddiman, D. C. *Analytical Chemistry* **2001**, 73, 3305-3311.

(50) Flora, J. W.; Muddiman, D. C. *Journal of the American Chemical Society* **2002**, 124, 6546-6547.

(51) Flora, J. W.; Muddiman, D. C. *Journal of the American Society of Mass Spectrometry* **2004**, 15, 121-127.

(52) Jones, J. L.; Dongre, A. R.; Somogyi, A.; Wysocki, V. H. *Journal of the American Chemical Society* **1994**, 116, 8368-8369.

(53) Guillaume, E.; Panchaurd, A.; Affolter, M.; Desvergnes, V.; Kussman, M. *Proteomics* **2006**, 6, 2338-2349.

(54) Lee, Y. H.; Kim, M.-S.; Choie, W.-S.; Min, H.-K.; Lee, S.-W. *Proteomics* **2004**, 4, 1684-1694.

(55) Bunk, D. M.; Macfarlane, R. D. International *Journal of Mass Spectrometry and Ion Processes* **1993**, 126, 123-136.

(56) Beardsley, R. L.; Karty, J. A.; Reilly, J. P. *Rapid Communications in Mass Spectrometry* **2000**, 14, 2147-2153.

Chapter 4

Implementing Photodissociation in an Orbitrap Mass Spectrometer

4.1 Overview

A dual pressure linear ion trap Orbitrap was modified to permit infrared multiphoton dissociation (IRMPD) in the higher energy collisional dissociation (HCD) cell for high resolution analysis. A number of parameters, including the pressures of the C-trap and HCD cell, the RF amplitude applied to the C-trap, and the HCD DC offset, were evaluated to optimize IRMPD efficiency and maintain a high signal-to-noise ratio. IRMPD was utilized for characterization of phosphopeptides, supercharged peptides, and N-terminal modified peptides, as well as for top-down protein analysis. The high resolution and high mass accuracy capabilities of the Orbitrap analyzer facilitated confident assignment of product ions arising from IRMPD.

4.2 Introduction

Over the past decade, both electron- and photon-based activation methods have emerged as versatile alternatives to collision-activated dissociation for fragmentation of peptides in proteomics applications.[1-5] Both infrared and ultraviolet photodissociation have proven successful tools for proteomic analysis using quadruple ion traps (QITs),[1] time-of-flight mass spectrometers,[6, 7] and Fourier transform ion cyclotron resonance (FT ICR) instruments;[8, 9] in the latter the high resolving power and mass accuracy have been particularly valuable for assignment of sequence ions.

Implementation of photodissociation on other high performance mass spectrometers is a compelling objective, especially with the growing adoption of Q-TOF[10] and Orbitrap[11, 12] type instruments for high throughput proteomics applications. Here we present the modification of an Orbitrap mass spectrometer to allow infrared multiphoton dissociation (IRMPD). We demonstrate a variety of applications that highlight the utility of IRMPD in combination with high resolution and mass accuracy Orbitrap mass analysis.

4.3 EXPERIMENTAL

4.3.1 Derivatization and Sample Preparation

PPITC derivatized peptides were prepared using 20 µL of a 10 µg/µL solution of reagent solution buffered to a pH of 9.5 with 20 mM $NaHCO_3$. It was incubated with 10 µL of 1 mM peptide solution, 60 µL 100 mM $NaHCO_3$ (pH 8 – 9) and 70 µL water for 30 minutes at 55º C. The product was purified on a C18 spin column and diluted to 10 µM prior to ESI analysis.

Supercharged peptides were prepared by mixing peptides at 10 µM in 49.5/49.5/1 methanol/water/acetic acid with 1% *m*-nitrobenzyl alcohol.

4.3.2 Mass Spectrometry

Samples were diluted to 10 µM in 49.5/49.5/1 water/methanol/acetic acid solution (v/v/v) and were run on a Thermo Scientific (San Jose, Ca) ETD-enabled LTQ Orbitrap XL that had been modified to allow entrance of a laser beam into the HCD cell as described below. Samples were directly infused via pico tips (New Objective, Woburn, MA). Both MS^1 and MS^2 scans were acquired at a resolving power of 60,000.

4.4 RESULTS AND DISCUSSION

4.4.1 Optimization of parameters and instrumentation set-up

IRMPD was performed in the HCD cell of a modified hybrid mass spectrometer comprising both linear ion trap and orbitrap mass analyzers (**Figure 4.1**).

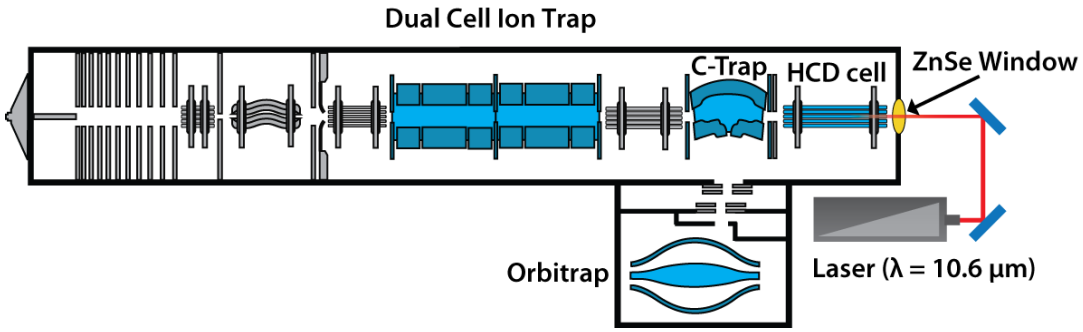

Figure 4.1 Schematic showing implementation of IRMPD in the HCD cell of a dual-cell linear ion trap-orbitrap

Typically utilized for beam-type CAD, the HCD cell provides a convenient region for photoexcitation of ions that have been mass-selected in the forward linear ion trap and transferred to the HCD cell.[13] Briefly, the HCD manifold was modified *via* the addition of a ZnSe window co-axial with the HCD cell and opening an aperture (d = ~2.54 mm) in its end plate electrode, allowing the irradiation of the HCD trapping region with IR photons (see **Figure 4.2**).

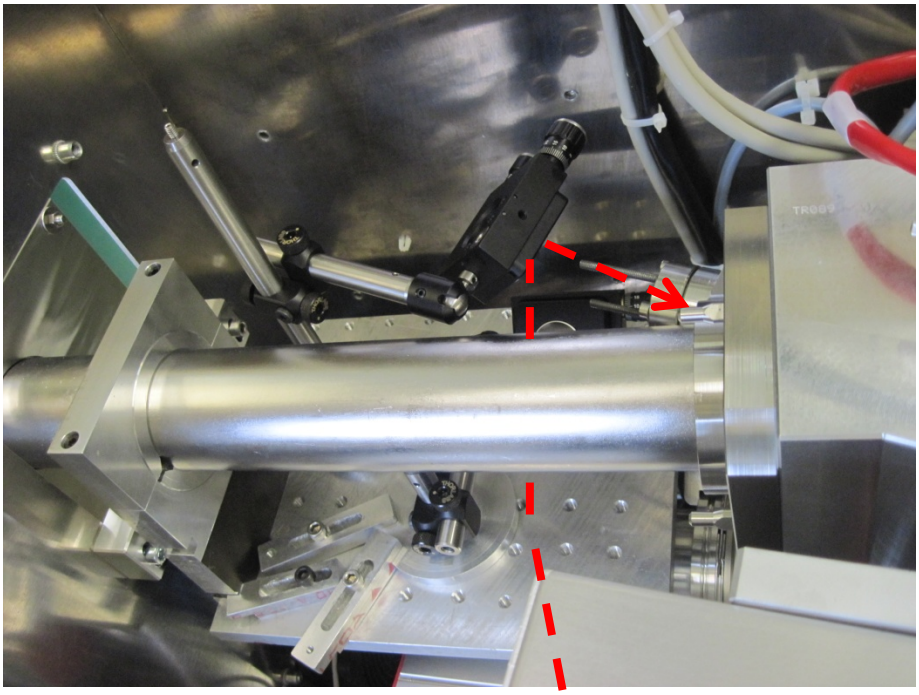

Figure 4.2 Introduction of laser into HCD cell; laser directed into HCD manifold through addition of a ZnSe window concentric with the HCD cell and two mirrors mounted through the use of a toe clamp.

Additionally, a nitrogen gas line was installed in the C-trap (a region normally used for spatial compression and temporally compact transfer of the ion bundle into the Orbitrap), and the top and bottom of the C-trap were fitted with ceramic plates to decrease gas conductance out of the C-trap, reducing the gas flow needed to achieve desired trapping efficiencies in the C-trap while minimizing the N_2 partial pressure in the QLT (quadrupole linear trap). Taken as a whole, these modifications allowed substantially independent control of nitrogen gas flow to both the C-trap and HCD cell. A 100 W Synrad CO_2 laser was directed in the HCD cell using two optic mirrors mounted behind

the HCD cell as shown in **Figure 4.2**. The beam width of the laser is 2.2 mm with a beam divergence of 7 mR.

The optimization of nitrogen flow to the HCD cell and C-trap is important for a number of reasons. IRMPD efficiency typically diminishes at higher pressures due to greater competition from collisional cooling.[14, 15] The trapping efficiencies of both the C-trap and HCD cells are pressure-dependent; with improved trapping efficiencies occurring at higher pressures, particularly for high *m/z* ions. To evaluate the overall effects of gas flow (*i.e.*, pressure) to both the HCD cell and C-trap on photodissociation efficiency, we evaluated three flow settings, one with reduced gas flow to both the C-trap and the HCD cell (P1), a second with a slightly increased flow to the C-trap while maintaining the reduced flow to the HCD cell (P2), and a third with no flow to the HCD cell and with the same increased flow to the C-trap as P2 (P3). Each flow setting led to corresponding changes in the pressures of the C-trap and HCD cell, and we observed notable changes in IRMPD signal levels and photodissociation efficiencies (**Figure 4.3**).

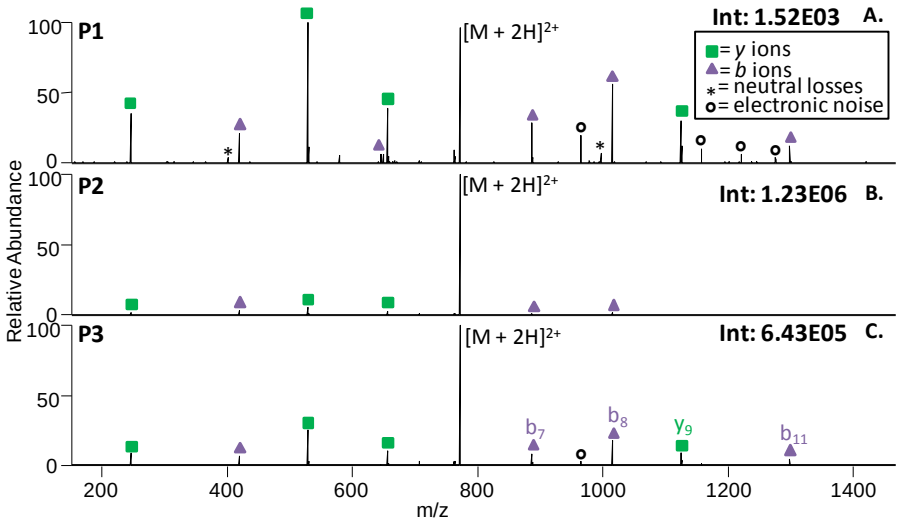

Figure 4.3 Effect of pressure on photodissociation efficiency and signal intensity. IRMPD of doubly charged phosphorylated peptide TSTEPQpYQPGENL, $[M + 2H]^{2+}$, (10 µM in 49.5/49.5/1 methanol/water/acetic acid), following 1 ms of 48 W irradiation at P1 (A), P2 (B), and P3 (C). RF amplitude = 1180 V_{p-p} (corresponding to first *m/z* = 100), collision energy = -11 V.

The P1 setting led to the lowest absolute signal levels, nearly one order of magnitude lower than those observed using the P2 setting, but with the highest photodissociation efficiency (*i.e.*, conversion of precursor ions to product ions). For the P2 setting, the photodissociation efficiency was relatively low due to collisional cooling, but the absolute signal level for the precursor ion was the highest. The P3 setting represented a compromise in the gas flow (and resulting pressures) that led to satisfactory precursor-to-product conversion efficiency and best product ion sensitivity, and therefore was used for all subsequent experiments.

In addition to pressure, the C-trap RF amplitude and HCD cell DC offset can influence overall IRMPD performance. As one parameter that defines the low-mass cut off, the C-trap RF amplitude is set as a function of the lower limit for the mass range selected through the user interface (also termed "first *m/z")*. The C-trap RF amplitude also influences the efficiency with which product ions are trapped prior to injection into the Orbitrap for mass analysis. Optimization of this parameter was needed in response to the pressure changes in the C-trap and HCD cell. As the C-trap RF amplitude varied from 600-2300 V_{p-p} (*i.e.*, corresponding to a first *m/z* range of 50 – 200 Th), the signal-to-noise (S/N) ratio increased substantially until leveling off at RF amplitudes above 1800 V_{p-p} (**Figure 4.4**). S/N was recorded for the most abundant fragment ion upon photoirradiation of a series of peptide precursor ions, and the averaged results are reported in **Figure 4.4**. To allow trapping of as many of the low *m/*z product ions as possible during IRMPD

while still maintaining a high S/N ratio, a C-trap RF amplitude of 1180 $V_{p\text{-}p}$ was used for most of the remaining experiments. We varied the HCD cell RF amplitude from 100 – 500 $V_{p\text{-}p}$, but observed no significant changes in photodissociation efficiency (data not shown), suggesting that, over this range of RF amplitudes, the ion cloud maintains a diameter smaller that the laser beam diameter.

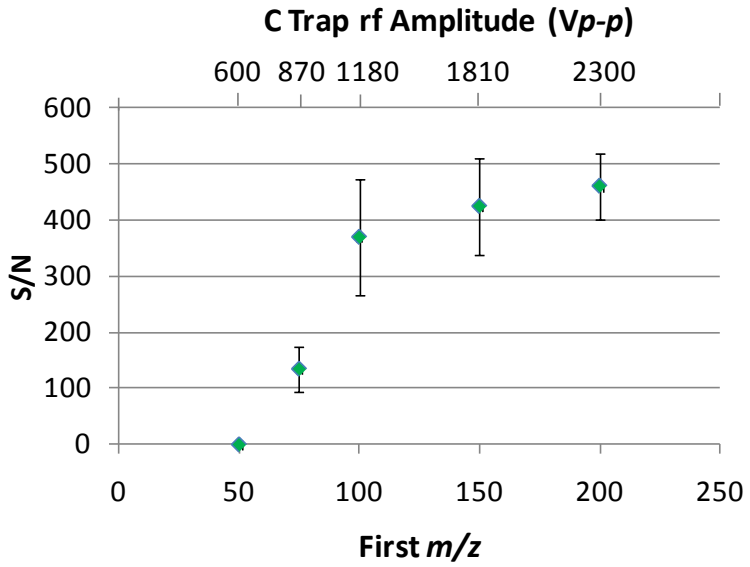

Figure 4.4 Variation of S/N with C-trap RF amplitude. S/N was measured for the most abundant fragment ion arising from several different peptide precursor ions as the C-trap RF amplitude was varied.

Additionally, the HCD cell DC offset, the potential difference between the C-trap and HCD cell, plays a significant role in IRMPD performance as a result of the reduced pressures in the HCD cell. The HCD cell DC offset determines the kinetic energy with which ions are transferred from the C-trap to the HCD region. Under normal instrument operating conditions, this parameter modulates the collision energy during beam-type

CAD, with greater HCD cell DC offsets providing greater collisional activation. At the reduced HCD cell pressures needed for efficient IRMPD, trapping of ions in the HCD cell is less effective. The HCD cell DC offset is normalized relative to precursor mass; as a result, no single HCD cell DC offset was optimal for the transfer of precursor ions. For the mass range of precursor ions utilized in this IRMPD study, HCD cell DC offsets of -8V to -11V relative to the QLT provided the best results in terms of the S/N ratio of isolated precursor (results not shown). For HCD cell DC offsets greater than -8V, the ions were not efficiently removed from the C-trap, and at offsets less than -11V, ions underwent collisional activation upon injection to the HCD cell. At HCD cell DC offsets of -8V to -11V, collisional activation prior to IRMPD was avoided while still efficiently extracting ions from the C-trap.

4.4.2 Applications

After optimization of the parameters summarized above, we utilized IRMPD in the hybrid Orbitrap system for a number of applications. Although conventional peptide cations exhibit relatively low IRMPD efficiencies due to significant competition with collisional cooling, phosphopeptides display ample IRMPD efficiencies even at mTorr pressures due to the high photoabsorptivities of P-O bonds at 10.6 μm.[16, 17] Phosphorylated peptides can be fully characterized and distinguished in a complex mixture with the phosphorylation site being easily pinpointed through IRMPD following short irradiation times.[18] As shown in **Figure 4.3C**, the phosphorylated peptide, TSTEPpYQPGENL, yielded an array of diagnostic sequence ions after just 1 ms of irradiation. Production of b_7, b_8, b_{11} and y_9 sequence ions, all of which retained the phosphoryl group, allowed facile localization of the site of phosphorylation. In addition

to phosphorylated peptides, peptides specifically derivatized *via* attachment of an IR chromophore can likewise be dissociated with high efficiency. As we described previously, peptides can be derivatized at their N-terminus using a phosphono derivatization agent, PPITC, which adds a strong IR chromophore.[19] Upon IRMPD, PPITC derivatized peptides dissociated into *b* and *y* ions with preferential cleavage at the y_{n-1} bond, as previously reported, using as little as 5-10 ms of IR irradiation (see **Figure 4.5**).

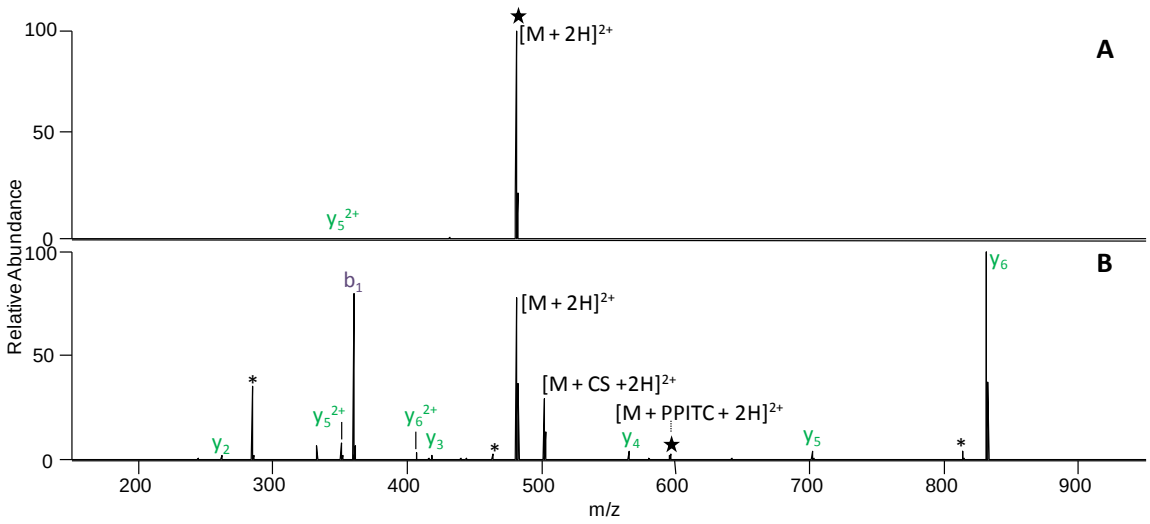

Figure 4.5 IRMPD of unmodified (**A**) MEHFRWG (2+) and (**B**) PPITC-derivatized MEHFRWG (2+) (10 µM in 49.5/49.5/1 methanol/water/acetic acid) following 10 ms of 48 W irradiation, C-trap rf amplitude = 1180 $V_{p\text{-}p}$, (corresponding to first *m/z* = 100), collision energy = -11V, * = neutral losses such as H_2O and NH_3, ★ = precursor, $[M + CS + 2H]^{2+}$ represents the ion arising after partial loss of the derivatization reagent.

Ions in higher charge states undergo photodissociation more efficiently than those in lower charge states,[20] an outcome largely attributed to the greater number of mobile

protons in highly charged peptide ions. The abundances of ions in higher charge states can be enhanced via a "supercharging" method in which a small amount of a solvent additive, like *m*-nitrobenzyl alcohol, alters the charging dynamics of ESI droplets and the resulting ions.[21, 22] This supercharging method also proved to be beneficial for improving the IRMPD efficiencies, as illustrated in **Figure 4.3** for supercharged peptide, ADSGEGDFLAEGGGVR, in the 3+ charge state.

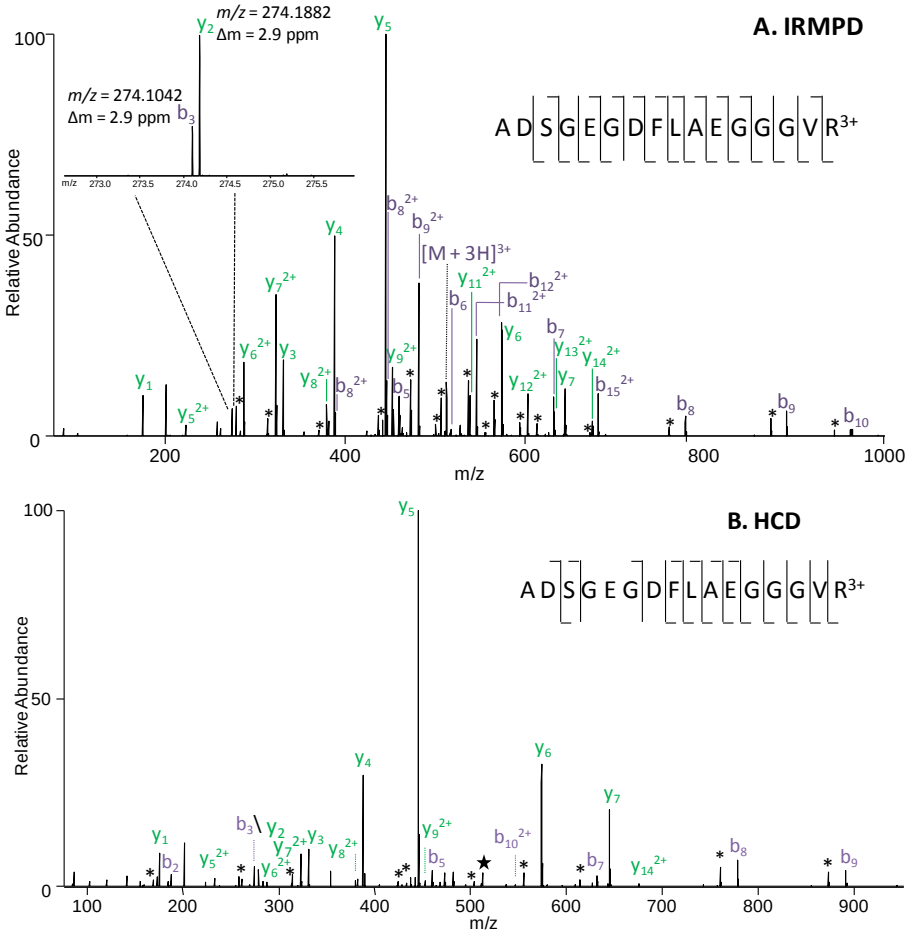

Figure 4.6 MS/MS of supercharged fibrinopeptide A (3+, sequence ADSGEGDFLAEGGGVR, 10 µM in 49.5/49.5/1 methanol/water/acetic

acid with 1% *m*-nitrobenzyl alcohol) (A) IRMPD using 8 ms of 48 W irradiation. * = neutral losses such as H_2O and NH_3, C-trap RF amplitude = 1180 $V_{p\text{-}p}$ (corresponding to first *m/z* = 100), collision energy = -11V, and (B) HCD DC offset ~ -18 V and C-trap rf amplitude = 870 $V_{p\text{-}p}$ (corresponding to first *m/z* = 75).

A large array of *b* and *y* ions was generated after 8 ms of irradiation. The expected masses of the b_3 ion, 274.1042 Da, and y_2 ion, 274.1882 Da, (a difference of 306 ppm from each other and 2.9 ppm from the experimental values) make them indistinguishable in lower resolution ion traps. IRMPD in the HCD cell with subsequent mass analysis in the Orbitrap allowed unambiguous assignment of these ions. In some but not all cases IRMPD results in a greater array of fragment ions than observed upon HCD, presumably due to consecutive IRMPD of primary fragment ions.

The enhanced resolution and mass accuracy is particularly beneficial for assignment of sequence ions upon IRMPD of whole proteins. **Figure 4.7** shows the IRMPD spectrum of the 12+ charge state of ubiquitin in which many of the fragment ions are highly charged, ranging from a singly charged b_2 ion to the multi-charged y_{70}^{11+} ion. With the increased resolution afforded by Orbitrap mass analysis, all photodissociation product ions were confidently distinguished and assigned. For example, in the inset of **Figure 4.7**, the y_{58}^{10+} ion was baseline-resolved with 3.5 ppm mass accuracy.

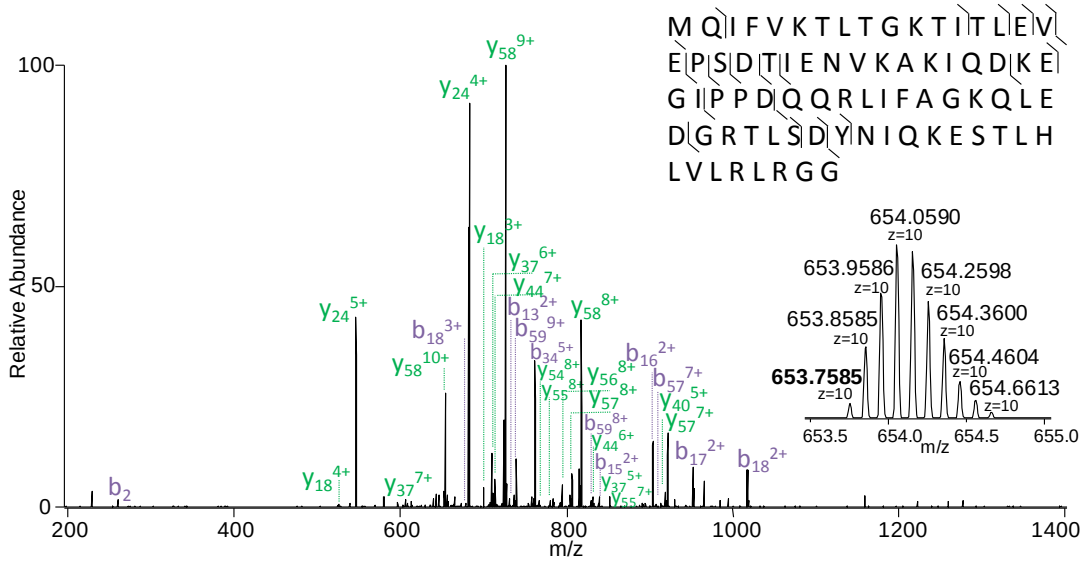

Figure 4.7 IRMPD of the 12+ charged state of ubiquitin (10 µM in 49.5/49.5/1 methanol/water/acetic acid) following 25 ms of 48 W irradiation, C-trap RF amplitude = 2300$V_{p\text{-}p}$ (corresponding to first *m/z* = 200), collision energy = -11V. Inset shows baseline resolution of the y_{58}^{10+} ion with a mass accuracy of 3.5 ppm relative to the theoretical *m/z* of 653.7562.

Interestingly, IRMPD also resulted in a greater degree of backbone cleavage selectivity compared to that obtained upon HCD with particular enhancement N-terminal to proline and C-terminal to glutamic acid residues. This phenomenon has been noted previously upon application of IRMPD in a dual pressure linear ion trap,[23] albeit without high resolution or high mass accuracy ion assignments. A comparison of cleavage selectivity obtained upon IRMPD versus HCD for the 10+, 11+, and 12+ charge states of ubiquitin is shown in **Figure 4.8**.

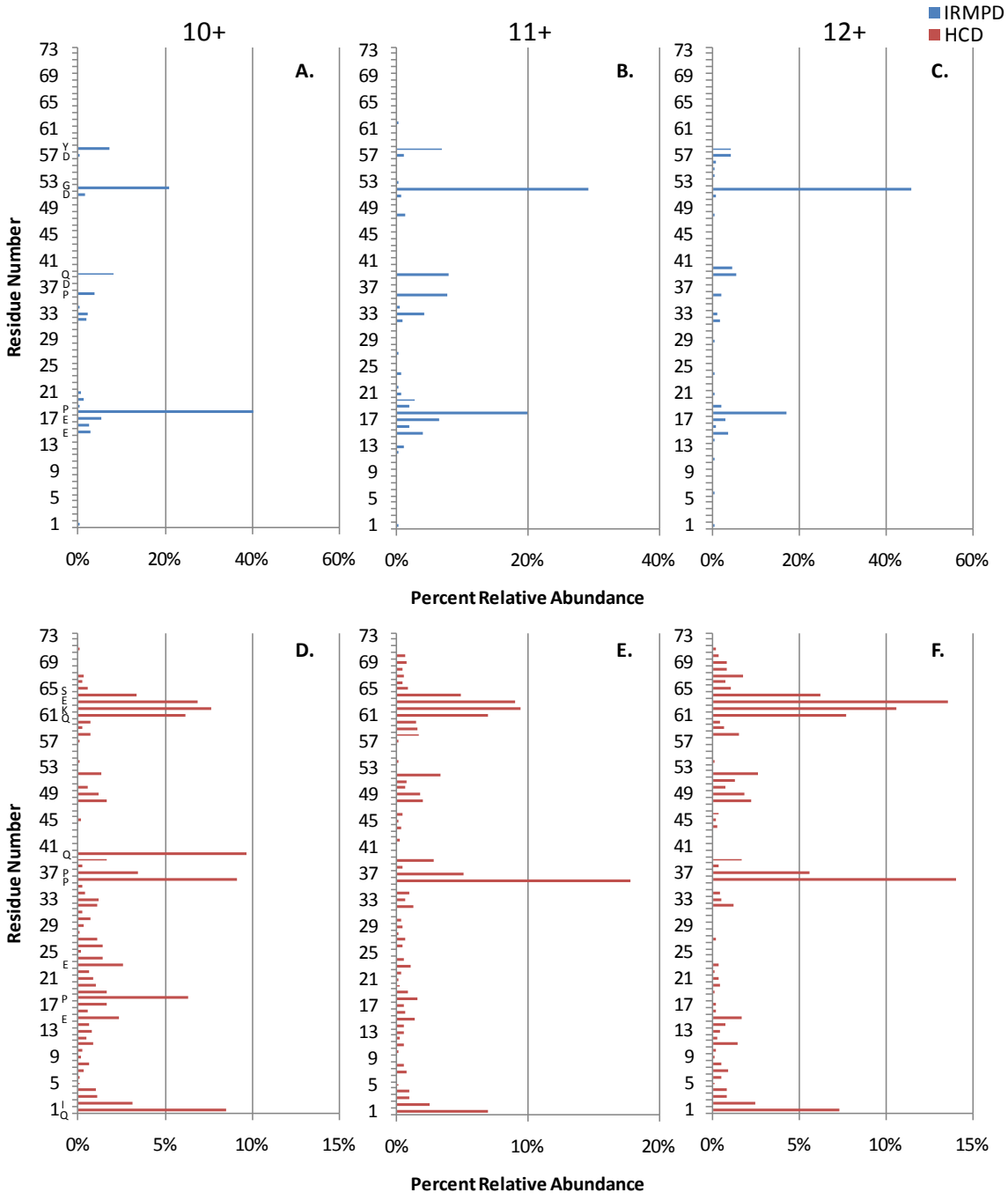

Figure 4.8 Summed b and y abundances for complementary ion pairs as a function of residue number from resulting cleavages of ubiquitin with IRMPD for (A)

$[M + 10H]^{10+}$, (B) $[M + 11H]^{11+}$, and (C) $[M+12H]^{12+}$; and with HCD for the (D) $[M + 10H]^{10+}$, (E) $[M + 11H]^{11+}$, and (F) $[M+12H]^{12+}$ ions.

4.5 CONCLUSIONS

The results summarized here illustrate the practical rewards of photodissociation in an Orbitrap hybrid mass spectrometer, providing high resolution analysis on a variety of samples. Photodissociation was accomplished in the HCD cell in which a decrease in its gas flow (and pressure), along with a concomitant increase in the gas flow to the C-trap, most effectively balanced IRMPD efficiency and overall ion abundance (S/N). IRMPD does not out-perform HCD, but it promotes an equally or more diverse array of fragment ions in some cases and demonstrates significantly enhanced performance for ions containing suitable chromophores. Maximizing both photodissociation and sensitivity requires further optimization of many of the parameters discussed.

4.6 REFERENCES

(1) Brodbelt, J. S.; Wilson, J. J. *Mass Spectrometry Reviews* **2009**, *28*, 390-424.

(2) Wiesner, J.; Premsler, T.; Sickmann, A. *Proteomics* **2008**, *8*, 4466-4483.

(3) Zubarev, R. In *Principles of Mass Spectrometry Applied to Biomolecules*, 2006, pp 475-517.

(4) Ly, T.; Julian, R. R. *Angewandte Chemie, International Edition* **2009**, *48*, 7130-7137.

(5) Coon, J. J. *Analytical Chemistry* **2009**, *81*, 3208-3215.

(6) Zhang, L.; Reilly, J. P. *Analytical Chemistry* **2009**, *81*, 7829-7838.

(7) Reilly, J. P. *Mass Spectrometry Reviews* **2009**, *28*, 425-447.

(8) Mann, M.; Kelleher, N. L. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 18132-18138.

(9) Eyler, J. R. *Mass Spectrometry Reviews* **2009**, *28*, 448-467.

(10) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *Journal of Mass Spectrometry* **2001**, *36*, 849-865.

(11) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. *Journal of Mass Spectrometry* **2005**, *40*, 430-443.

(12) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Analytical Chemistry* **2006**, *78*, 2113-2120.

(13) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nature Methods* **2007**, *4*, 709-712.

(14) Black, D. M.; Payne, A. H.; Glish, G. L. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 932-938.

(15) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. *Analytical Chemistry* **2009**, *81*, 8109-8118.

(16) Flora, J. W.; Muddiman, D. C. *Analytical Chemistry* **2001**, *73*, 3305-3311.

(17) Flora, J. W.; Muddiman, D. C. *Journal of the American Chemical Society* **2002**, *124*, 6546-6547.

(18) Crowe, M. C.; Brodbelt, J. S. *Analytical Chemistry* **2005**, *77*, 5726-5734.

(19) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 377-384.

(20) Madsen, J. A.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 349-358.

(21) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. *Analytical Chemistry* **2001**, *73*, 1455-1460.

(22) Iavarone, A. T.; Williams, E. R. *International Journal of Mass Spectrometry* **2002**, *219*, 63-72.

(23) Madsen, J. A.; Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. *Analytical Chemistry* **2009**, *81*, 8677-8686.

Chapter 5

Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores

5.1 OVERVIEW

Two N-terminal derivatization reagents containing aromatic chromophores, 4-sulfophenyl isothiocyanate (SPITC) and 4-methylphosphonophenyl isothiocyanate (PPITC), were used to increase the dissociation efficiencies of peptides upon ultraviolet photodissociation (UVPD) at 193 nm. The resulting UVPD spectra are dominated by C-terminal ions, including *y*, *z*, *x*, *v*, and *w* ions, and immonium ions. The attachment of the PPITC or SPITC groups leads to a reduction in the number and abundances of N-terminal ions because the added phosphonate or sulfonate functionalities result in neutralization of some of the N-terminal species, ones that might normally be singly protonated in the absence of the negatively charged sulfonate or phosphonate groups. In addition, the greater photoabsorptivities of the PPITC- and SPITC-derivatized N-terminal product ions enhanced their secondary photodissociation, leading to formation of immonium ions.

5.2 INTRODUCTION

The ability to address increasingly complex questions in proteomics has motivated the development and application of a variety of new tandem mass spectrometric strategies over the past decade. Although collision induced dissociation[1] remains the most popular activation method, electron-based methods, including electron capture dissociation[2, 3] and electron transfer dissociation[4], and photon-based activation

(photodissociation)[5-7] have gained recognition as promising alternatives. In particular, photodissociation affords fast, tunable energy deposition based on the wavelength used and the irradiation time. Ultraviolet photodissociation has proven well-suited for proteomic applications due to the high UV photoabsorptivity of peptide amide bonds.[8] Irradiation with high energy UV photons promotes a collection of high energy fragments, primarily *a*, *x* and side chain loss (*d*, *v*, and *w*) ions in time-of-flight mass spectrometers while *b* and *y* fragments also remain abundant in ion trap instruments.[9] In addition, UVPD can be implemented on the nanosecond to microsecond timescale by using fast pulsed lasers, providing incentive for its adaptation to high throughput applications. It has been noted that some amino acid side chains, specifically the aromatic residues,[10-12] play a more important role in the absorption of UV radiation. It was reported that upon UVPD at 266 nm the tryptophan (W), tyrosine (Y), or phenylalanine (F) residues in a peptide led to a chromophore effect in which cleavage near the aromatic residue was increased and lower irradiation times were needed to significantly deplete the parent ion population. These previous studies showcased the impact of specific chromogenic functionalities on UVPD behavior and motivated our interest in exploiting this factor for enhancing UVPD efficiencies in proteomic applications via derivatization of peptides to incorporate aromatic moieties.

There have been a number of derivatization strategies implemented for proteomic applications, including those designed to alter the fragmentation patterns of peptides[13-16] and others developed to attach specific chromophores.[17-23] For example, N-terminal sulfonation has proven to be a useful tactic for enhancing *de novo* sequencing of peptides via generation of simplified MS/MS patterns.[13, 17, 28] Through the incorporation of a

negatively charged group at the N-terminus of a peptide, all singly charged N-terminal fragments are neutralized which suppresses the presence of *b*-type or other N-terminal fragment ions in the resulting MS/MS spectra. This progression towards enhancement of a single ion type, more specifically the C-terminal ions in this case, facilitated the interpretation of MS/MS product ion spectra using *de novo* sequencing methods. A variation of this method utilized 4-methylphosphonophenylisothiocyanate (PPITC) to attach an IR-chromogenic phosphonite group to the N-terminus of peptides which increased their IRMPD efficiencies and improved their individual sequence coverage.[17] This type of derivatization strategy was also incorporated into a UVPD method in which UV chromophores were attached at the N-termini of peptides via an AlexaFluor sulfonation reagent, endowing the peptides with high photoabsorptivity at 355 nm and simplifying the resulting spectra to a single series of *y* ions.[20] As another example of an innovative derivatization method, Julian and coworkers has pursued site-selective dissociation of peptides upon UVPD at 266 nm by modification of cysteine residues with quinone moieties. The integrated cysteine modification/UVPD method resulted in backbone cleavage at the modified cysteine site.[22] These prior studies have demonstrated the ability to enhance and/or tune the dissociation properties of molecules based on incorporation of specific chromophores, an approach that we extend in the present work.

In this study, two N-terminal derivatization reagents, 4-sulfophenyl isothiocyanate (SPITC) and 4-methylphosphonophenyl isothiocyanate (PPITC),[17] were utilized to increase the dissociation efficiencies of peptides upon UVPD at 193 nm via attachment of aromatic groups that serve as UV chromophores (see **Figure 5.1**). The activation and dissociation behavior of the resulting SPITC- and PPITC-derivatized peptides were

compared to that of the underivatized peptides to demonstrate the enhanced UVPD upon addition of simple aromatic chromophores. Moreover, the ionizable PPITC and SPITC functional groups reduce the number and abundances of N-terminal ions for both singly and multiply charged precursor ions, and thus the resulting UVPD spectra are heavily dominated by C-terminal ions including *y*, *z*, *x*, *v*, and *w* ions, as well as immonium ions. This N-terminal derivatization/UVPD strategy allows confident identification of the entire peptide sequence including the N-terminal residue due to the formation of abundant b_1 and a_1 ions.

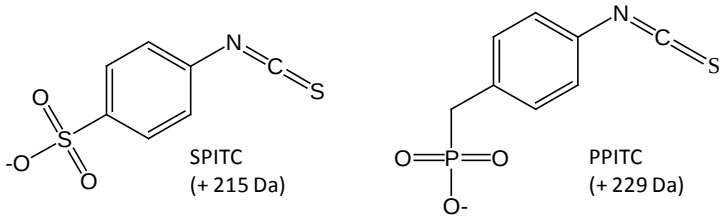

Figure 5.1 Structures of N-terminal derivatization reagents (shown as deprotonated forms); mass shifts associated with each are indicated in parenthesis.

5.3 EXPERIMENTAL

5.3.1 Materials and Reagents

All chemicals were purchased from Sigma Aldrich (St. Louis, MO) except for the following peptides which were obtained from BACHEM (King of Prussia, PA): FSWGAEGQR, DRVYIHPFHL, DAEFRHDSGYQVHHQK, ASHLGLAR, YRPPGFSPFR, GNHWAVGHLM, MEHFRWG, and RQpSVELHSPQSLPR. ADSGEGDFLAEGGGVR was obtained from the American Peptide Co. (Sunnyvale, CA).

5.3.2 Derivatization and Sample Preparation

For all model peptides, 20 μL of a 10 μg/ μL reagent solution buffered to pH 8.5 with 20 mM $NaHCO_3$ was incubated with 10 μL of 1 mM peptide solution for 30 minutes at 55 ^{o}C. For bovine serum albumin (BSA), the protein was reduced by mixing 11 μL of 150 mM dithiothreitol with 40 μL of 1 mM BSA for 45 minutes at 55 ^{o}C, followed immediately by alkylation with the addition of 12 μL of 550 mM iodoacetamide (IAM) buffered to a pH of ~8.0 with NH_3HCO_3) and reacting for 30 minutes in the dark at room temperature. The protein was diluted with 100 mM NH_4HCO_3 and digested with 1 mg/mL trypsin in 1 mM HCl in a 1:100 w/w ratio overnight at 37 ^{o}C. The reduced and alkylated protein was then desalted using a C18 spin column and reacted with each reagent as described above. Reactions were all terminated with a small addition of 1.5 μL 5% trifluoroacetic acid followed by desalting on a C18 spin column. The eluent was diluted to 10 μM before ESI analysis with 49.5/49.5/1 H_2O/MeOH/Acetic Acid.

5.3.3 Mass Spectrometry and Liquid Chromatography

All experiments were performed on a ThermoFinnigan LTQ XL linear ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) equipped with a standard ESI source. The LTQ mass spectrometer was modified for UVPD as described earlier[24] with a CaF_2 optical window to allow introduction of photons from a Coherent ExciStar XS laser operated at 193 nm (Coherent Lambda Physik, Germany). The laser was directed through a 1.7 mm entrance aperture that was in line with the 2 mm aperture of the exit lens of the ion trap. The typical laser pulse energy was 8 mJ operating at 500 Hz.

Solutions were directly infused at a flow rate of 3 μL/min at a concentration of 10 μM in 49.5:49.5:1 MeOH:H_2O:Acetic acid. An ESI voltage of 4 kV and a heated capillary temperature of 180^{o} C was used for all mass spectrometry experiments. All UVPD experiments were performed using a q_z value of 0.1 with the total activation period ranging from 0.03 msec (the minimum default activation period for a single 5 ns laser pulse using Thermo Excalibur instrument software) to 10 msec (sufficient for 5 laser pulses). Photodissociation efficiencies were calculated as follows:

$$PD\ Efficiency\ (\%) = \left[\frac{\left(\sum Area\ of\ \ product\ ions \right)}{\sum Area\ of\ \ product\ ions + Area\ of\ \ surviving\ \ precursor} \right] \times 100$$

(Equation 5.1)

Liquid chromatography was performed using a Dionex UltiMate 3000 system (Sunnyvale, CA). An Agilent ZORBAX 300SB-C_{18} column (Santa Clara, CA) (150 × 0.3 mm, 5 μm particle size) was used for all separations. The column temperature was kept constant throughout the run at 40°C. Eluent A consisted of 0.1% formic acid in water and eluent B 0.1% formic acid in acetonitrile. A linear gradient from 5% eluent B to 40% eluent B over 65 min at 5 μL/min was used. Injections of approximately 1 μg (20 picomoles) were used for the digested BSA sample. For all LC-MS/MS runs, the first event was the full mass scan (*m/z* range of 400 – 2000) followed by ten consecutive MS/MS (either UVPD or CID) events on the ten most abundant ions from the full mass scan. A *q*-value of 0.1 and one, 5 ns UV pulse applied per MS/MS scan was used for UVPD activation, while a *q*-value of 0.25, an activation time of 30 ms, and a normalized collision energy (NCE) of 35% were used for CID events. The maximum injection time for events

was set to 100 ms, and each centroid mass spectrum and tandem mass spectrum was the average of three microscans. For all LC-UVPD experiments a background subtraction was also utilized due to a high abundance of photoionization products seen around and below *m/z* 200. These ions have been hypothesized to be the products of photoionization of organics species that reside in the vacuum system.[25] Background spectra were collected in the centroid mode while no sample was being infused so that no ions were being trapped and using an isolation window of *m/z* 500 ± 4. The average of 20 scans was then subtracted from the LC-UVPD .RAW file using the Thermo Fisher Xcalibur version 2.0.7 software.

5.4 RESULTS AND DISCUSSION

5.4.1 UVPD of Model Peptides

UV photodissociation has many attributes that make it extremely appealing for proteomic applications, including its ability to produce a variety of sequence ions (depending on the wavelength) using very short activation times.[5, 7] The large array of sequence ion types make the resulting UVPD spectra amenable to database search algorithms, as demonstrated in our recent investigation of the applicability of UVPD for identification of proteins from a human cell lysate using an activation of a single 5 ns pulse of 193 nm irradiation.[26] However, we noticed that the UV photodissociation efficiencies were lower for tryptic peptides that did not possess any aromatic residues, in some cases hampering sequence identification. In fact, aromatic residues have been shown to greatly increase the absorptivity of peptides at 266 nm.[10-12] We sought to uniformly increase the photodissociation efficiencies of peptides at 193 nm by attaching suitable chromophores, in this case aromatic moieties. We have previously utilized two

N-terminal derivatization schemes and demonstrated that in combination with IRMPD[17] or ETD[27] these reactions substantially improved the *de novo* sequencing of peptides, primarily by eliminating the redundant *b* or *c* ion series that impede differentiation of the *b* and *y* ions or *c* and *z* ions. Here the same N-terminal derivatization reagents, SPITC and PPITC, are used to modify a series of peptides to specifically enhance their absorptivities and consequently their photodissociation efficiencies at 193 nm via the incorporation of aromatic chromophores. Although the SPITC and PPITC reagents have similar UV chromophores, the acidities of the sulfonate and phosphonate substituents are sufficiently different to warrant examination of both of these derivatization reagents and their impact on the formation and survival of diagnostic sequence ions in the present UVPD strategy. A third peptide derivatization agent, phenylisothiocyanate (PITC), was also considered because it appends the same aromatic chromophore. However, the rather low peptide derivatization efficiency of PITC compared to SPITC and PPITC renders it less viable for proteomic applications.[17]

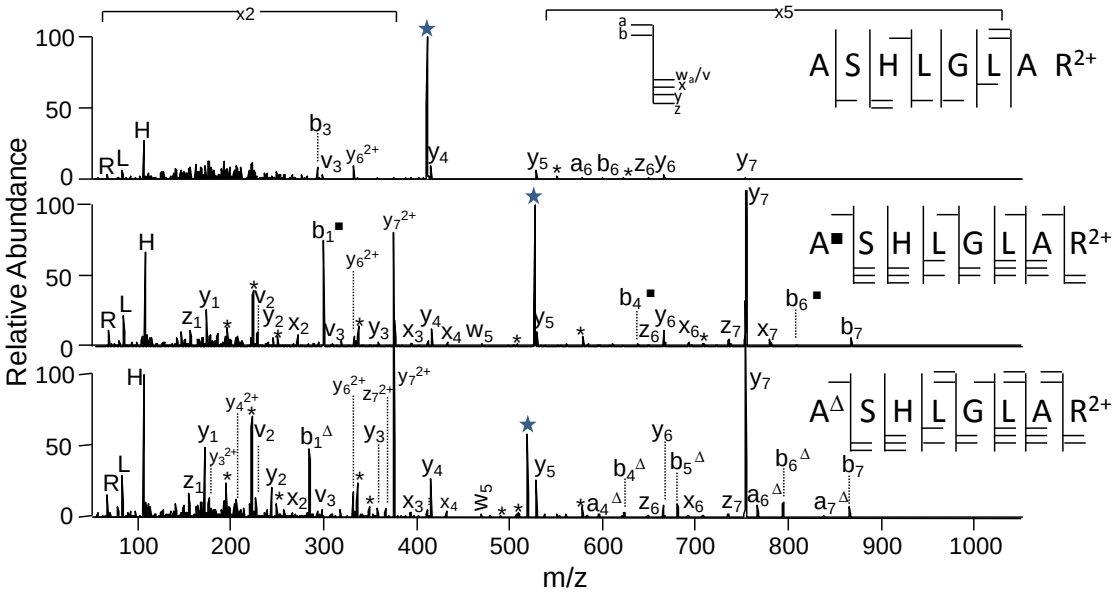

Figure 5.2 UVPD comparison of (A) underivatized, (B) SPITC-derivatized and (C) PPITC-derivatized ASHLGLAR (2+) after a single 8 mJ pulse of 193 nm irradiation. * represents internal ions, ■ = retains SPITC modification, Δ = retains PPITC modification, ★ represents the remaining precursor ion.

Examples of the resulting UVPD spectra are shown above in **Figure 5.2** for the doubly protonated peptide ASHLGLAR. This peptide has no aromatic side-chains in its amino acid sequence, and in its unmodified form displays a moderately low degree of dissociation, giving a photodissociation efficiency of ~30% after a single pulse of UV irradiation. An array of sequence ion types are produced, including *a*, *b*, *z*, *y*, and immonium ions, but the net sequence coverage is incomplete based on using a fragment ion abundance threshold of 0.5% relative to the precursor abundance. In addition, all of the fragment ions arising from backbone cleavages have rather low abundances, thus limiting the ability to correctly identify the peptide using standard algorithms and also reducing the overall sensitivity of the method. After derivatization using either the SPITC

or PPITC reagent, a single pulse of UV irradiation was sufficient to produce abundant sequence ions (**Figure 5.2B** and **5.2C**, respectively). Of particular high abundance is the $y_{(n-1)}$ peak that represents the Edman degradation type cleavage that is commonly seen after N-terminal PITC-type derivatization in which the N-terminal amino acid is most readily cleaved.[15, 17, 27, 28] This cleavage near the chromophore arises from a fragmentation pathway facilitated by the PITC moiety and is observed upon CID or UVPD. An array of other sequence ions, including *a*, *b*, *x*, *y*, *z*, and immonium ions are also produced. The b_1 product ion is highly abundant despite the fact that it contains the aromatic group with its associated sulfonate or phosphonate moiety, indicating that the latter must remain protonated in the b_1 product ions.

The photodissociation efficiencies (as defined in equation 1 in the experimental section) were measured for a series of peptides to evaluate the impact of the aromatic chromophores. As summarized in **Figure 5.3**, the non-derivatized peptides exhibit photodissociation efficiencies ranging from 28% to 66% (average is 50%), with the efficiencies generally increasing for those peptides with one or more amino acids containing aromatic side-chains. In comparison, the photodissociation efficiencies range from 50% to 95% for the SPITC- and PPITC-derivatized peptides, with the efficiency increasing an average of 20% relative to the non-derivatized analogs.

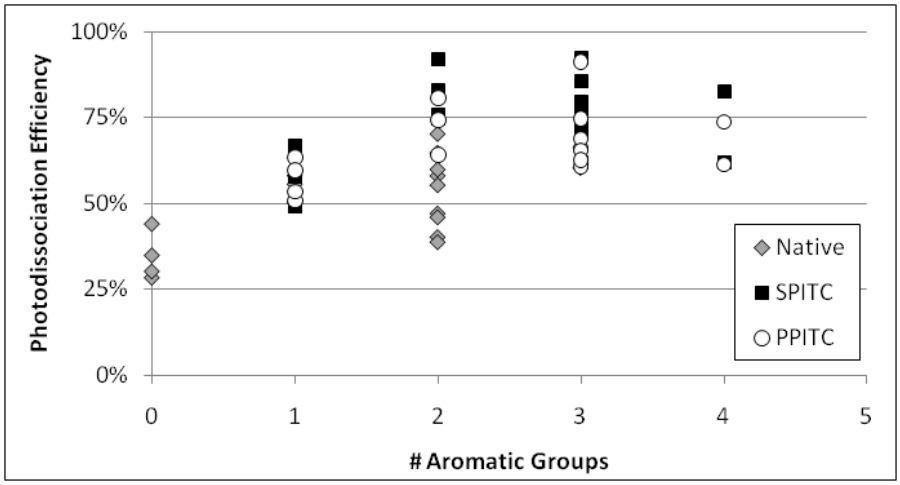

Figure 5.3 Photodissociation efficiencies obtained for UVPD of non-derivatized and N-terminally derivatized peptides after a single 8 mJ pulse of 193 nm irradiation.

In addition, the two PITC reagents offer the added benefit of simplifying the UVPD spectra through the neutralization of many N-terminal fragments due to the presence of the ionizable phosphonate or sulfonate groups. The distribution of C-terminal (*v*, *w*, *x*, *y* and *z*) and N-terminal (*a* and *b*) product ions was examined for a series of eight peptides both in their unmodified and PPITC- or SPITC-derivatized forms, and the results are summarized in **Figure 5.4** for each accessible charge state of each peptide.

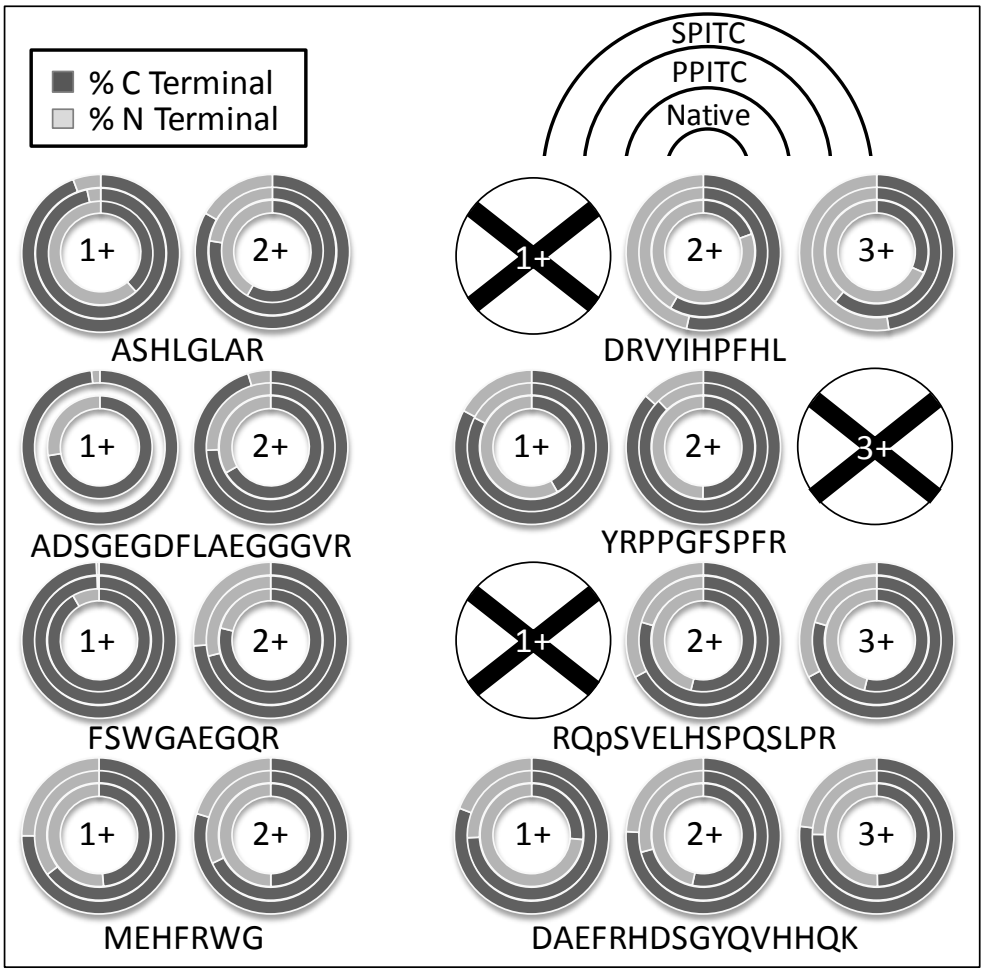

Figure 5.4 Distribution of C-and N-terminal Ions Produced by UVPD (single pulse, 193 nm) of Peptides.[a]

[a] The m/z value of the singly-charged PPITC-ADSGEGDFLAEGGGVR1 falls above the m/z range of the mass spectrometer and thus could not be analyzed. An "x" indicates when a charge state was not present for a specific peptide, thus preventing data acquisition.

The portion of N-terminal ions is uniformly lower for the PITC-derivatized peptides than for the unmodified peptides, with the difference being particularly striking for a few of the peptides, such as ASHLGLAR and YRPPGFSPFR. As expected, the elimination of

N-terminal fragment ions is more notable for the singly charged PITC-modified peptides because of the lower probability that the N-terminal portion of the peptide will end up sequestering two protons (one localized at the ionizable PITC group and one at the amine terminus) to maintain a net charge of 1+. The singly charged PITC-peptides yielded the most simplified MS/MS spectra that could be utilized in *de novo* sequencing applications which exploit specific series of ions for prediction of peptide sequences. With respect to the UVPD spectra of the PITC-derivatized peptides, the formation of *w* and *v* (C-terminal) ions is likewise advantageous because these side-chain loss fragment ions allow distinction between isobaric amino acid residues such as isoleucine and leucine that pose particular challenges for identification by conventional activation methods like CID. **Figure 5.2** shows an example of this additional benefit upon inspection of the *w* and *v* ions. In **Figure 5.2A**, the v_3 ion confirms that the third amino acid from the C-terminus is Leu; however, with only a y_5 ion available to identify the fifth amino acid from the C-terminus, differentiation between Leu and Ile is not possible. In contrast, in both **Figure 5.2B** and **5.2C** the presence of the w_5 ion allows unambiguous identification of the fifth amino acid as Leu based on the side chain loss.

Interestingly, the multiply charged peptide ions also showed a significant shift in the distribution of C-terminal versus N-terminal ions upon derivatization as seen in **Figure 5.4**. The enhancement of C-terminal ions and neutralization of N-terminal ions (due to the presence of the ionizable chromophore) would be expected to be most pronounced for singly charged precursors which have a limited number of mobile protons. The continuation of the significant preference for C-terminal ions for the multiply charged PITC-peptides may be rationalized by several factors. First, the C-

terminal Edman degradation type $y_{(n-1)}$ ions remain dominant products even for the multi-charged peptides (as demonstrated also upon CID of the PITC-derivatized peptides, data not shown). Second, even for CID or UVPD of multi-charged peptides, many of the fragments are singly charged, suggesting that net neutralization of N-terminal products containing an ionizable group remains probable. Moreover, it is well known that product ions can absorb one or more additional photons and undergo further fragmentation (i.e. secondary dissociation). We have demonstrated this concept with UVPD at 355 nm utilizing the N-terminal derivatization chromophore, AlexaFluor 350 in which the chromophore-containing N-terminal ions absorbed additional photons and dissociated.[20] In contrast, the C-terminal ions did not contain UV chromophores and remained as stable product ions, thus resulting in MS/MS spectra dominated by C-terminal ions.[20] We speculate that the N-terminal products ions formed upon UVPD in the present study, all which contain the attached SPITC or PPITC aromatic chromophore, have high UV absorption efficiencies that lead to their efficient secondary photodissociation. Although the data summarized in **Figure 5.5** was collected using only a single laser pulse, it is conceivable that first generation product ions may absorb a second photon during the first UV pulse and dissociate into non-detectable products. This process would eliminate or reduce their contribution to the resulting UVPD spectra.

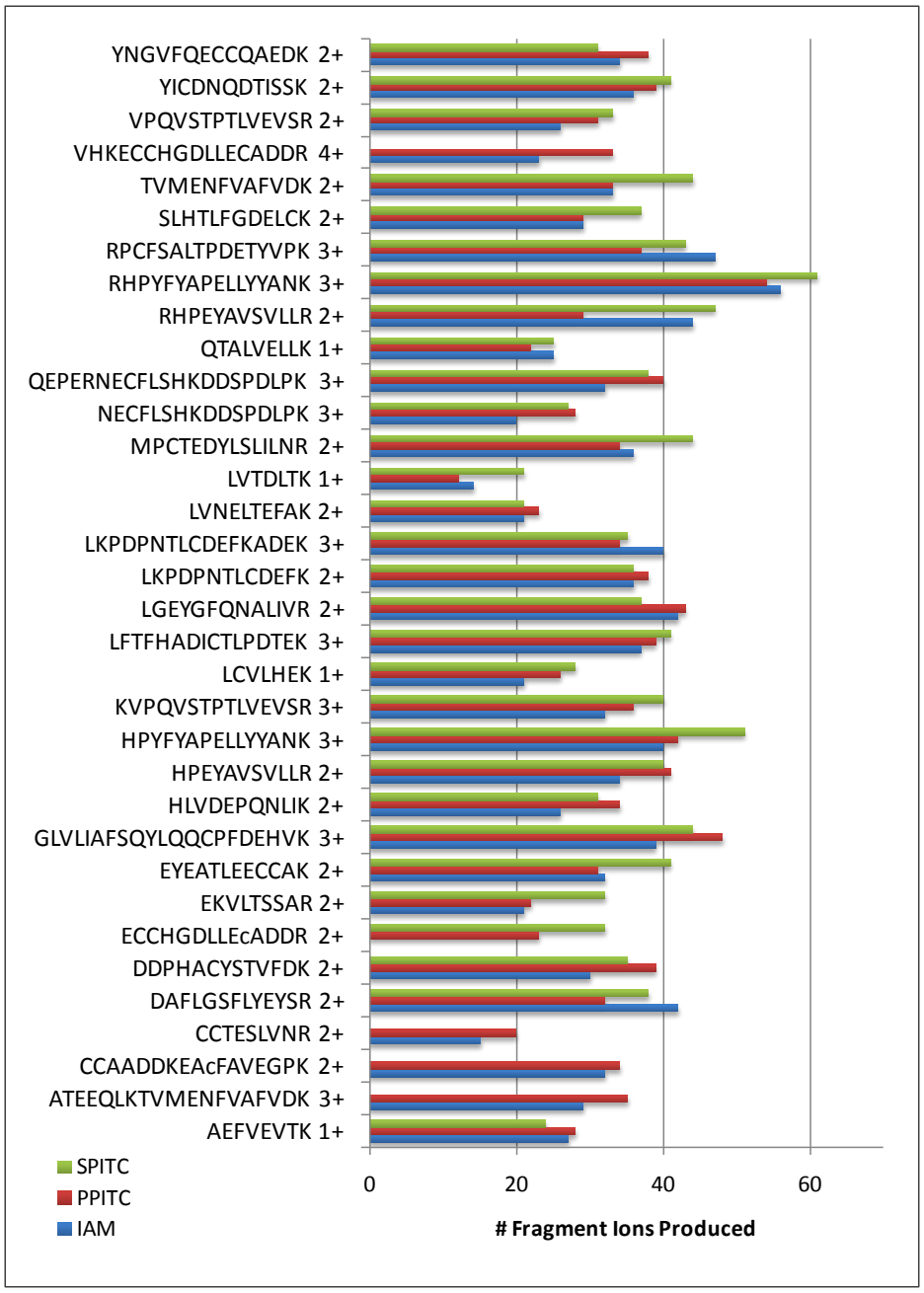

Figure 5.5 Total number of fragment ions produced for a tryptic digest of BSA after a single 8 mJ pulse at 193 nm.[a]

[a]Some of the SPITC-peptides did not show up in any of the LC runs and thus could not be included.

The most abundant N-terminal ions for the SPITC and PPITC-peptides upon UVPD are the b_1 and a_1 ions, as noted earlier. Although the relative abundances of the C-terminal ions increased in comparison to the N-terminal ions with the number of pulses, in general the dominant product ions became the immonium ions, thus confirming that both the N-terminal and C-terminal sequence ions dissociate into simple immonium ions. This secondary fragmentation into predominant immonium ions also occurred for the underivatized peptides as well and demonstrates the fact that multiple pulses of irradiation may not be an effective option for increasing the portion of backbone fragments (i.e. N-terminal and C-terminal sequence ions) produced upon UV photodissociation. In all cases, the greatest relative abundances of sequence ions were generated with a single pulse of irradiation. Phenylisothiocyanate (PITC) was also evaluated as a UV chromophore. While PITC-derivatized peptides yielded similar UVPD results as those derivatized by SPITC or PPITC, the poor peptide reaction efficiency (20-40%) compared to the other two (50-90%) rendered it less promising for protein applications.

5.4.2 UVPD of BSA Tryptic Digest

The same derivatization reactions were applied to a tryptic digest of BSA in order to evaluate the level of improvement in UVPD efficiencies for a more complex mixture. For this assessment, bovine serum albumin was reduced, alkylated and digested, divided into three aliquots with one unmodified and the other two subjected to SPITC or PPITC derivatization, then analyzed by LC-UVPD-MS. As shown in **Figure 5.6**, the number of sequence ions (including *x*, *y*, *z*, *v*, *w*, *a*, *b*, and *c*) produced following a single 8 mJ pulse

of UV irradiation is summarized for all peptides identified. Although the total number of diagnostic ions remains similar for the PITC-derivatized and non-derivatized tryptic peptides, the impact on sequence coverage and peptide identification is rather substantial because of the change in the distribution of N-terminal and C-terminal ions, as described earlier for the model peptides.

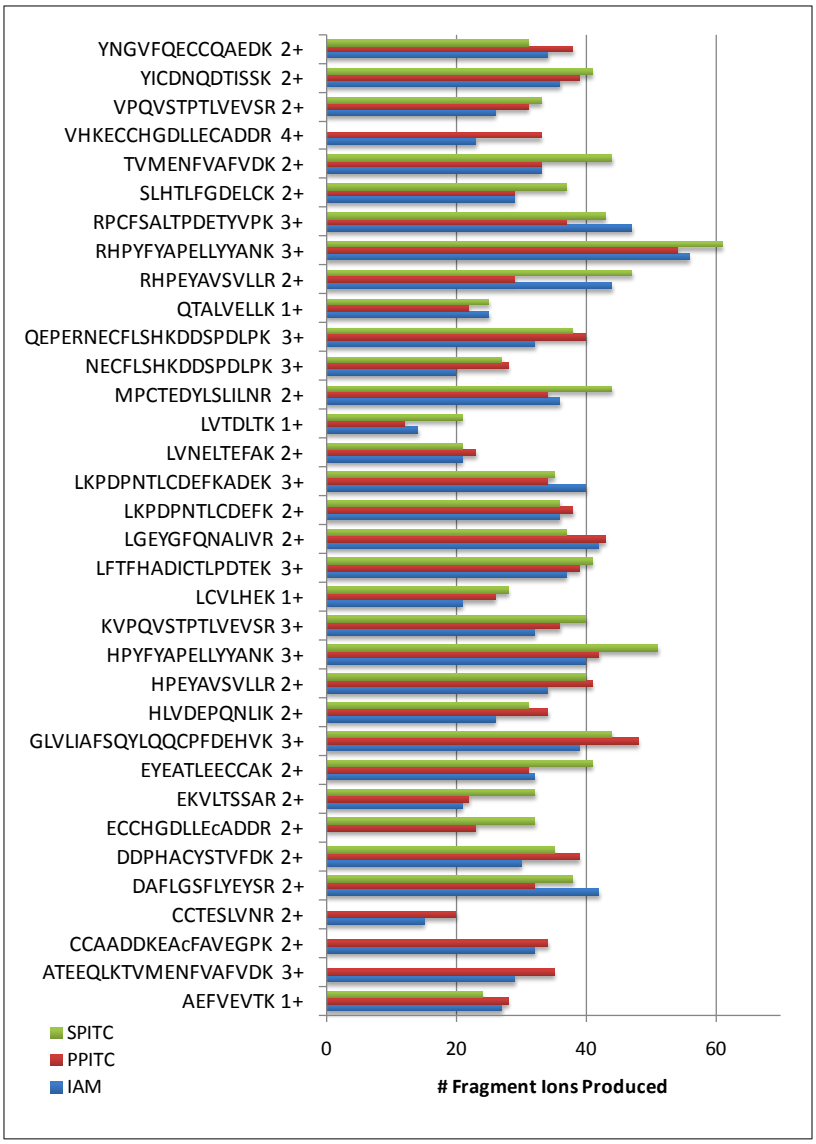

Figure 5.6 Total number of fragment ions produced for a tryptic digest of BSA after a single 8 mJ pulse at 193 nm.[a]

[a]Some of the SPITC-peptides did not show up in any of the LC runs and thus could not be included.

For the PITC-derivatized peptides, the number of N-terminal (*a*, *b*, *c*) ions decreases whereas the number of C-terminal (*x*, *y*, *z*, *v*, *w*) ions correspondingly increases. Moreover, the photodissociation efficiencies are greater for the PITC-derivatized peptides (see **Figure 5.7**), especially ones that are produced in 1+ or 2+ charge states. The impact of an additional aromatic group is especially noticeable for the tryptic peptides containing zero to one aromatic amino acid. The increase in dissociation efficiency leads to more abundant fragment ions which increase the overall sensitivity of the method. This improvement in dissociation efficiency is beneficial when utilizing database search algorithms in which the abundances of key sequence ions must occur above a certain signal-to-noise threshold (typically three) to ensure the successful identification of peptides.

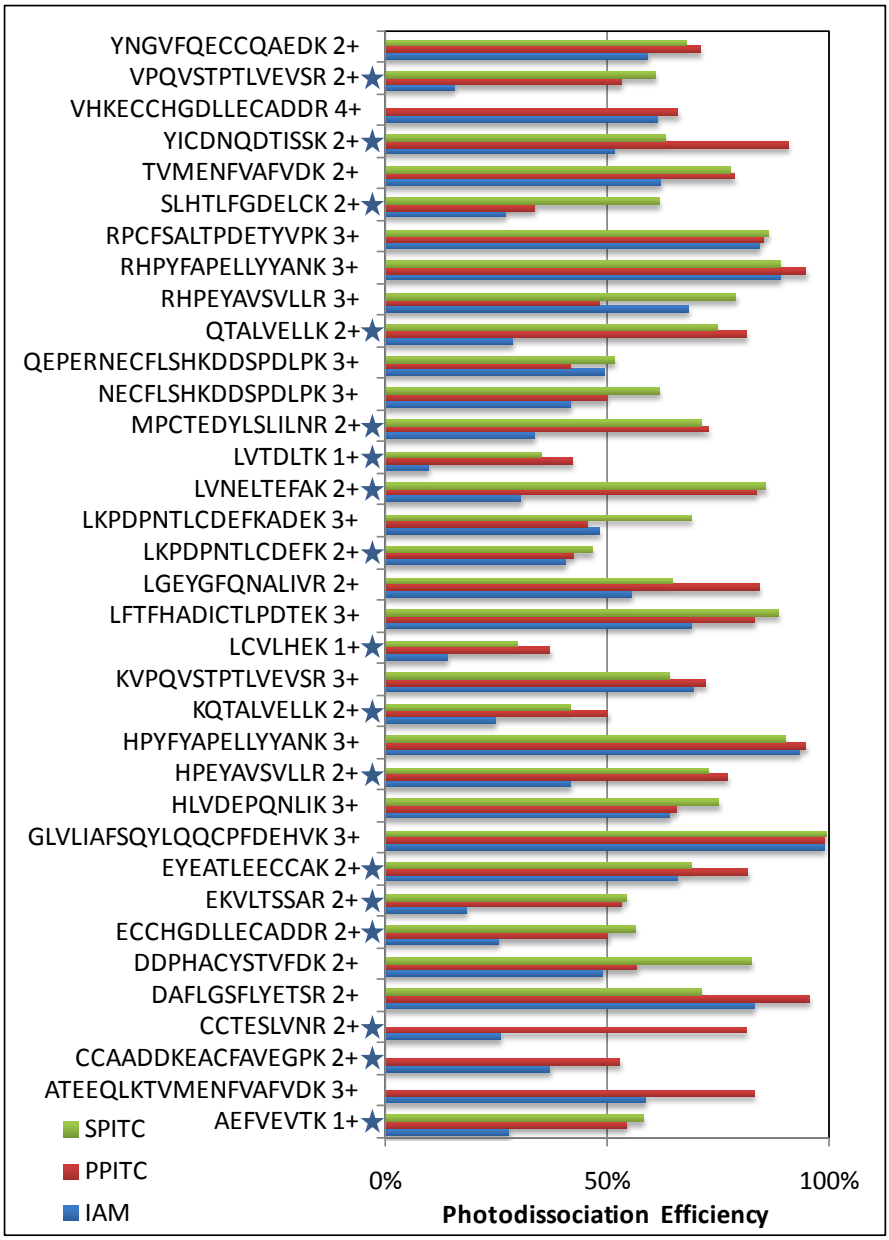

Figure 5.7 UV photodissociation efficiencies of tryptic BSA peptides after a single 8 mJ pulse at 193 nm.[b]

[b]The ★ designates those peptides with a single or no aromatic amino acids and that generate predominantly ions in the 1+ to 2+ charge state. Some of the SPITC-peptides did not show up in any of the LC runs which accounts for the missing green bars.

The bar graphs in Figure **5.8** display the relative distributions of N-terminal and C-terminal products ions for the PITC-modified peptides compared to the analogous alkylated peptides (IAM-alkylated). The abundances of all of the C-terminal or N-terminal sequence ions were summed and used to create product ion distribution maps, with the exclusion of the N-terminal b_1 and a_1 ions whose abundances were exaggerated due to the preferential cleavage of the N-terminal amino acid for the PITC-peptides. The alkyated but non-chromophore-containing IAM-peptides show a higher abundance of N-terminal ions, averaging 37%, whereas the SPITC- and PPITC-peptides averaged 11% N-terminal ions and 89% C-terminal ions. For the SPITC- and PPITC-peptides, the a_1 and b_1 ions actually constitute the largest percentage of N-terminal ions upon UVPD of peptides in charge states greater than one, while most of the other N-terminal ions are much lower in abundance if present at all. The IAM-peptides produced fewer C-terminal ions that collectively averaged 63% of the distribution.

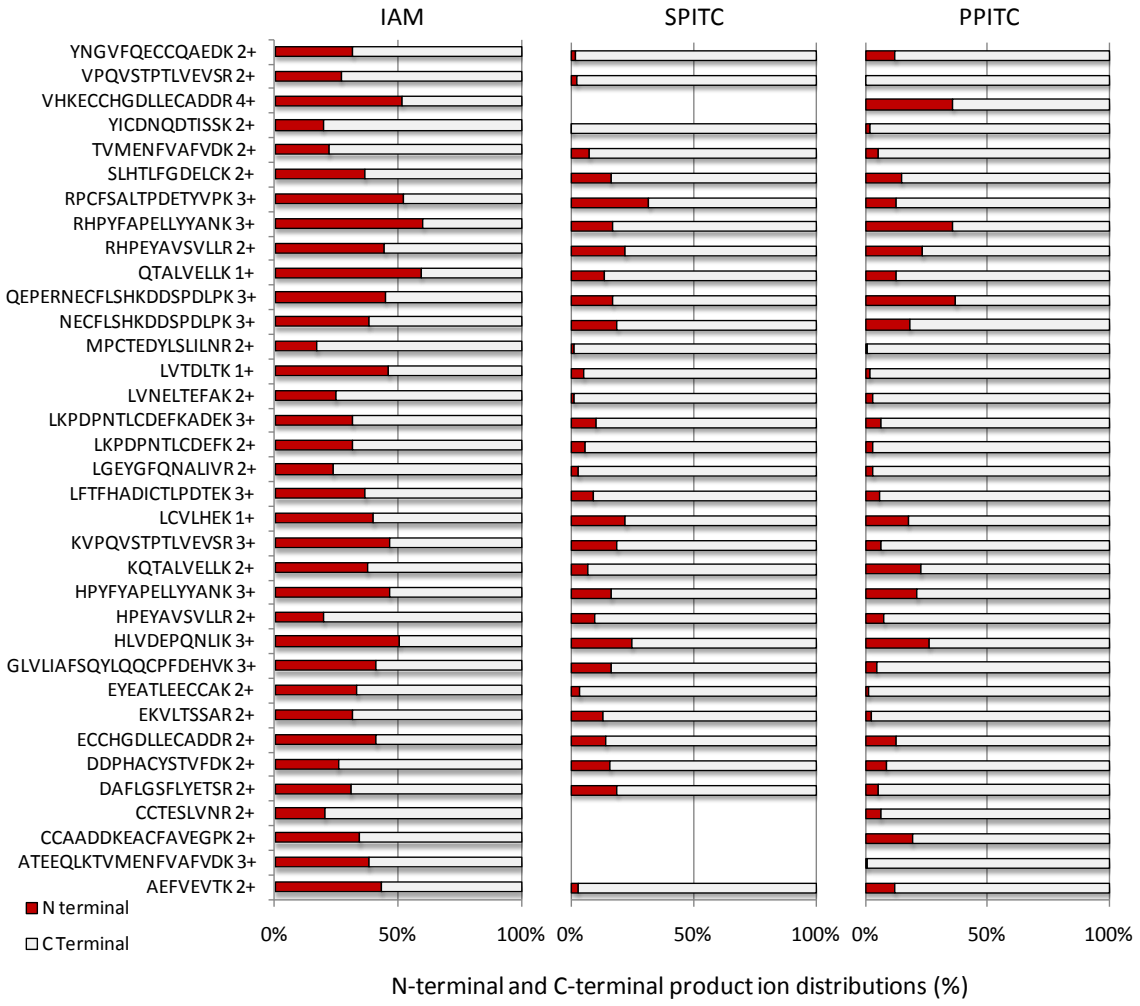

Figure 5.8 N-terminal and C-terminal product ion distributions for BSA tryptic peptides based on the type of derivatization method, following a single 8 mJ pulse of UV irradiation in which the relative portion of N-terminal (red bars) and C-terminal (white bars) sequence ions is calculated by (Σ Abundance of all N or C Terminal Ions[c])/(Σ Abundances of all C and N ions).

[c]For the N-terminal data, all N-terminal ions are included except the b_1 and a_1 ions. Some of the SPITC-peptides did not show up in any of the LC runs which accounts for the missing bars. IAM means iodoacetamide alkylation.

5.5 CONCLUSIONS

Two N-terminal peptide derivatization reagents containing aromatic chromophores were utilized to enhance photodissociation of peptides at 193 nm. The attachment of aromatic groups increased the photodissociation efficiencies by up to a factor of two and resulted in production of a slightly greater number of diagnostic fragment ions. While the overall increase in the number of sequence ions was not dramatic, there was a significant increase in the number and abundances of C-terminal ions and corresponding decrease in the number and abundances of N-terminal ions. The number and abundances of N-terminal ions were reduced due to two factors. The presence of the ionizable PPITC or SPITC functionalities resulted in neutralization of some of the N-terminal species (ones that might normally be singly protonated without the negatively charged sulfonate or phosphonate group). Moreover, the higher photoabsorptivities of the N-terminal PITC-modified product ions enhanced their secondary dissociation upon absorption of additional photons, resulting in fragmentation into immonium ions or elimination of redundant N-terminal product ions from the spectra.

5.6 REFERENCES

(1) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212-217.

(2) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563-573.

(3) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.

(4) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528-9533.

(5) Reilly, J. P. *Mass Spectrom. Rev.* **2009**, *28*, 425-447.

(6) Brodbelt, J. S.; Wilson, J. J. *Mass Spectrom. Rev.* **2009**, *28*, 390-424.

(7) Ly, T.; Julian, R. R. *Angew. Chem., Int. Ed.* **2009**, *48*, 7130-7137.

(8) Peterson, D. L.; Simpson, W. T. *J. Am. Chem. Soc.* **1957**, *79*, 2375-2382.

(9) Thompson, M. S.; Cui, W.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1439-1452.

(10) Oh, J. Y.; Moon, J. H.; Kim, M. S. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2706-2712.

(11) Oh, J. Y.; Moon, J. H.; Kim, M. S. *J. Mass Spectrom.* **2005**, *40*, 899-907.

(12) Oh, J. Y.; Moon, J. H.; Lee, Y. H.; Hyung, S.-W.; Lee, S.-W.; Kim, M. S. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1283-1288.

(13) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 7131-7136.

(14) Wang, S.; Zhang, X.; Regnier, F. E. *J. Chromatogr., A* **2002**, *949*, 153-162.

(15) Lee, Y. H.; Shin, J.-W.; Ryu, S.; Lee, S.-W.; Lee, C. H.; Lee, K. *Anal. Chim. Acta* **2006**, *556*, 140-144.

(16) Leon, I. R.; Neves-Ferreira, A. G. C.; Valente, R. H.; Mota, E. M.; Lenzi, H. L.; Perales, J. *J. Mass Spectrom.* **2007**, *42*, 1363-1374.

(17) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 377-384.

(18) Wilson, J. J.; Brodbelt, J. S. *Anal Chem* **2006**, *78*, 6855-6862.

(19) Pikulski, M.; Wilson, J. J.; Aguilar, A.; Brodbelt, J. S. *Anal. Chem.* **2006**, *78*, 8512-8517.

(20) Wilson, J. J.; Brodbelt, J. S. *Anal. Chem.* **2007**, *79*, 7883-7892.

(21) Diedrich, J. K.; Julian, R. R. *J. Am. Chem. Soc.* **2008**, *130*, 12212-12213.

(22) Diedrich, J. K.; Julian, R. R. *Anal. Chem.* **2010**, *82*, 4006-4014.

(23) Ly, T.; Julian, R. R. *J. Am. Chem. Soc.* **2008**, *130*, 351-358.

(24) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80*, 4807-4819.

(25) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1657-1665.

(26) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.*, ACS ASAP.

(27) Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2009**, *81*, 3645-3653.

(28) Wang, D.; Kalb Suzanne, R.; Cotter Robert, J. *Rapid Commun Mass Spectrom* **2004**, *18*, 96-102.

Chapter 6

Enhanced Electron Transfer Dissociation through Fixed Charge Derivatization of Cysteines

6.1 OVERVIEW

Electron transfer dissociation (ETD) has proven to be a promising new ion activation method for proteomics applications due to its ability to generate *c*- and *z*-type fragment ions in comparison to the *y*- and *b*-type ions produced upon the more conventional collisional activation of peptides. However, low precursor charge states hinder the success of electron-based activation methods due to competition from non-dissociative charge reduction and incomplete sequence coverage. In the present report the reduction and alkylation of disulfide bonds prior to ETD analysis is evaluated by comparison of three derivatization reagents: iodoacetamide (IAM), N,N-dimethyl-2-chloro-ethylamine (DML), and (3-acrylamidopropyl)-trimethyl ammonium chloride (APTA). While both the DML and APTA modifications lead to an increase in the charge states of peptides, the APTA-peptides provided the most significant improvement in percent fragmentation and sequence coverage for all peptides upon ETD, including formation of diagnostic ions that allow characterization of both the C- and N-termini. In addition, the formation of product ions in multiple charge states upon ETD is minimized for the APTA-modified peptides.

6.2 INTRODUCTION

There have been tremendous inroads in methodologies for the activation and dissociation of ions for proteomics applications in recent years. Although collision induced dissociation (CID) remains one of the most popular activation methods, largely due to a well-developed understanding of peptide fragmentation patterns,[1-3] other methods that address some of the more challenging sequencing problems have gained momentum. For example, due to its tunable and well-defined energy deposition, surface induced dissociation has proven to be very effective for peptide fragmentation, as well as more recently for examination of protein macromolecular complexes.[4] Photodissociation has also emerged as a promising alternative ion activation method for biological molecules, offering efficient energization of ions via a collision-free photoabsorption process.[5, 6] Most recently, the development of electron-based dissociation methods, including electron capture dissociation (ECD)[7, 8] and electron transfer dissociation (ETD),[9, 10] has further expanded the suite of ion activation strategies used for proteomic applications.

The electron-based activation methods entail reactions of low energy electrons (ECD) or radical anions (ETD) with multiply charged peptides or proteins, leading to preferential cleavages of the N-C_α bonds along the backbone that result in the production of complementary *c*- and $z^{\bullet}$-type ions. The peptide length and amino acid composition have shown little influence on the fragmentation patterns of these dissociation methods in large scale studies[11] although two more detailed targeted studies have reported some dependence.[12, 13] ECD and ETD have proven to be particularly well-suited for the

identification of post-translational modifications of proteins as these labile modifications are retained on the amino acid side chains during ion activation and dissociation, and thus can be readily tracked, whereas they are often lost with other ion activation methods. Despite their tremendous promise, the electron-based dissociation methods exhibit a significant charge state dependence reflected by the fact that ions in lower charge states tend to undergo charge reduction without subsequent dissociation into informative *c* and $z^{\bullet}$ sequence ions.[8, 14, 15] The use of trypsin as a standard enzymatic reagent to cleave proteins prior to MS/MS analysis has in some ways exacerbated the tendency for charge reduction upon ECD or ETD because the large natural abundance of lysine and arginine residues in proteins results in the production of relatively short peptides upon tryptic digestion that yield lower charge states upon ESI.

Several methods have been aimed at alleviating this charge-state dependence of ETD and ECD, either utilizing auxiliary energization of ions to convert them to more meaningful sequence ions[14-20] or by increasing the charge states of the precursor ions prior to activation.[16-19] The former has been achieved by implementing supplemental collisional activation of the charge-reduced non-dissociated precursor ions or by irradiating the precursor ions with photons to disrupt the noncovalent bonds formed between *c*- and $z^{\bullet}$-type product ion pairs.[14-20] Elevating the temperature of the ions in an ICR cell via blackbody heating provided sufficient energy to the precursor ions to destroy the noncovalent bonding and increase the formation of *c*- and $z^{\bullet}$-type fragment ions upon ECD.[20, 21] In a process termed activated ion-ECD (AI-ECD) by McLafferty et al., simultaneous collisional activation and electron irradiation (i.e., ECD) of precursor ions also proved to be a successful approach for alleviating the extent of non-dissociative

charge reduction of precursor ions.[22, 23] Both the Coon and McLuckey groups have employed supplemental collisional activation of charge-reduced products from electron transfer dissociation reactions, improving peptide sequence coverage.[11, 24, 25] More recently, the Coon group implemented photoirradiation during ETD of peptides in a linear ion trap mass spectrometer in a method analogous to AI-ECD described above.[26] This method significantly improved the sequence coverage of peptides having higher precursor m/z values (i.e. lower charge states).

Another approach to alleviating the charge state dependence of ETD efficiency is to produce precursor ions in higher charge states either through super-charging or derivatization techniques. The addition of *m*-nitrobenzyl alcohol to the ESI solutions promotes super-charging of ions and subsequent ETD analysis of these highly charged precursor ions was enhanced through higher sequence coverage.[27] Another method was recently described for the analysis of peptides by derivatization of cysteine residues in peptides using N,N-dimethyl-2-chloro-ethylamine, which converts cysteines to highly basic dimethyl lysine (DML) moieties leading to a higher net charge states upon ESI and greater peptide sequence coverage upon ETD.[28] Alternatively, the analyte molecules of interest can be derivatized by attachment of fixed charge sites to shift them to higher charge states prior to electron-based activation.[16-19] McLuckey et al. incorporated fixed charges via attachment of trimethylammonium moieties to primary amine groups in disulfide-linked peptides and demonstrated that the relative preference for backbone cleavage or disulfide bond cleavage upon ETD was influenced by the number of mobile protons versus fixed charges.[18] Other groups have demonstrated the benefits of fixed

charge derivatization of peptides in conjunction with ECD. For example, Chamot-Rooke at al. showed that the addition of trimethyl-phenyl phosphonium cations to the N-termini of *O*-glycosylated and *O*-phosphorylated peptides increased the sequence coverage upon ECD.[16, 17] O'Connor et al. reported the derivatization of lysine side-chains and/or N-termini via attachment of trimethylpyridinium and noted an increase in side-chain cleavages with a corresponding decrease in backbone cleavages upon ECD.[19]

In this study we explore an alternative derivatization strategy based on the use of an S-alkylation reagent, (3-acrylamidopropyl)-trimethyl ammonium chloride (APTA), which introduces fixed charges to all free cysteine side-chains of peptides through a quaternary ammonium moiety, in conjunction with ETD analysis. The APTA reaction has been shown previously to improve the ionization efficiencies of peptides, as well as their water solubility, while also allowing the use of strong cation exchange chromatography for the separation of the derivatized peptides.[29, 30] Our primary interest was exploiting the addition of fixed charge sites to enhance the use of electron transfer dissociation for the characterization of cysteine-containing peptides. In this study, APTA is used to derivatize cysteine-containing peptides, including those from a tryptic digest of bovine serum albumin (BSA), prior to characterization by ETD-MS.

6.3 EXPERIMENTAL

6.3.1 Materials and Reagents

All reagents were purchased from Sigma Aldrich (St. Louis, MO, USA), and all peptides were obtained from BACHEM (King of Prussia, PA, USA), including

FQVVCG, HCKFWW, CDPGYIGSR, NRCSQGSCWN, somatostatin (AGCKNFFWKTFTSC), TGF-alpha (CHSGYVGVRC), and bactenecin (RLCRIVVIRVCR). Immobilized TPCK trypsin beads were purchased from Pierce Biotechnology (Rockford, IL, USA). Acetonitrile and methanol (HPLC grade) were purchased from Fisher Scientific (Fair Lawn, NJ, USA).

6.3.2 Derivatization and Sample Preparation

All disulfide bonds in the peptides or proteins were first reduced by adding a 5-fold molar excess of dithiothreitol (DTT) to the peptide or protein (10 μL, 1 mM) in aqueous solution and incubating at 40 °C for 1 hr in the dark (per manufacturer's recommended protocol for optimal efficiency). Immediately following reduction, the reduced peptide samples were split into three aliquots; one for iodoacetamide (IAM) alkylation, one for APTA derivatization, and one for DML modification. IAM-alkylation was performed by adding iodoacetamide (4 μL, 1 M) in ammonium bicarbonate buffer (pH ~8.0, 100 mM). This reaction was allowed to proceed at room temperature for 1 hr in the dark, followed by the addition of excess DTT to quench the alkylation reaction. The resulting products are ones in which the reduced thiols are converted to *S*-carboxamidomethyl groups. APTA derivatization was achieved by reacting the reduced peptide solution with APTA (16 μL, 1 mg/mL in 50:50 ACN/ 100 mM ammonium bicarbonate buffer, pH 8.0) overnight at room temperature in the dark followed by freezing to terminate the reaction. The resulting products are fixed-charge trimethylammonium *S*-alkyl cysteine derivatives. DML modifications were initiated by reacting N,N-dimethyl-2-chloro-ethylamine (1 μL, 1 mM in 1M HEPES buffer, pH 7.8)

for three hours at room temperature, producing carboxyamidomethyl cysteines. After incubation all peptides were desalted using Thermo Fisher PepClean C_{18} spin columns and diluted for MS analysis as described above.

For the BSA protein sample, 100 μL of immobilized TPCK-treated trypsin beads in 100 mM ammonium bicarbonate (pH ~8.0) were used to digest 10 μmol of protein. The protein was enzymatically digested overnight at 37 °C after which the beads were removed and the resulting tryptic digest was subjected to IAM-alkylation or APTA-derivatization using the procedures outlined above.

6.3.3 Mass Spectrometry

Samples were diluted to 10 μM in 49.5/49.5/1 water/methanol/acetic acid solution (v/v/v) and were directly infused at 3 μL/min into a Thermo Fisher LTQ XL (San Jose, CA, USA) with ETD capabilities. All ETD experiments were conducted with fluoranthene as the radical anion electron-transfer reagent with a 100 ms reaction time. A Dionex UltiMate 3000 capillary high performance liquid chromatography (Sunnyvale, CA, USA) was used for all separations. Samples with a final concentration of either 10 μM enzymatically digested protein or APTA-derivatized protein (1 μL injections) were introduced to a reversed-phase C_{18} column. Eluents consisted of 0.1% formic acid in water (A) and 0.1% formic in acetonitrile (B) and were used to generate a gradient of 5% B for 5 min followed by a linear gradient to 90% B over 45 min at a flow rate of 4 μL/min. Data dependent analysis was conducted for all LC-MS/MS analyses by first acquiring an ESI mass spectrum (*m/z* 400-2000) followed by CID and ETD analysis

undertaken on the two most abundant peaks. A normalized collision energy of 35% was used for all LC-CID-MS analyses.

6.3.4 Protein Identification

Protein samples were identified using Sequest search software using the following parameters: 2 maximum miscleavages, ± 0.5 Da fragment mass tolerance, ± 1.0 Da precursor mass tolerance and the appropriate fixed modification of Cys (either the iodoacetamide alkylation or the APTA alkylation).

6.4 RESULTS AND DISCUSSION

6.4.1 ETD of Model Peptides

It has been previously shown that the addition of a fixed charge to a peptide can lead to a significantly different fragmentation pattern from its native form.[17, 31-33] For example, fixed-charge derivatization schemes have been utilized extensively to simplify tandem mass spectra through the elimination of one series of fragment ions using a variety of different activation methods including PSD[32], CID[33], IRMPD[34], ECD[17], and recently in our group with electron transfer with gentle collision activation (ETcaD).[35] In the present study, we aim to improve the sequence coverage of peptides upon ETD analysis by increasing the charge states of peptides via attachment of quaternary amines to reduced cysteine residues. Cysteine residues are a logical choice for implementation of selective derivatization strategies owing to their high reactivity and low abundance in proteins which allows streamlined subsets of peptides to be targeted. As demonstrated in this study, the APTA derivatization procedure selectively and efficiently derivatizes all

reduced cysteine residues and leads to higher charge states that result in greater sequence coverage, including production of key diagnostic ions that allow characterization of both the C- and N-termini. We also compare the ETD results obtained for APTA-derivatized peptides having fixed charge sites to peptides in which the cysteines are reduced and IAM-alkylated and to the same peptides for which the reduced cysteine residues are converted to dimethyl lysine analogs, which are more basic moieties that also afford higher charge states (but not fixed charge sites). IAM is commonly used to prevent the recombination of disulfide bonds in reduced peptides by formation of S-carboxamidomethyl derivatives which are easily ionized by ESI and MALDI-MS. The DML derivatization procedure was shown previously to substantially increase both ETD efficiencies and sequence coverage obtained for cysteine-rich peptides.[22]

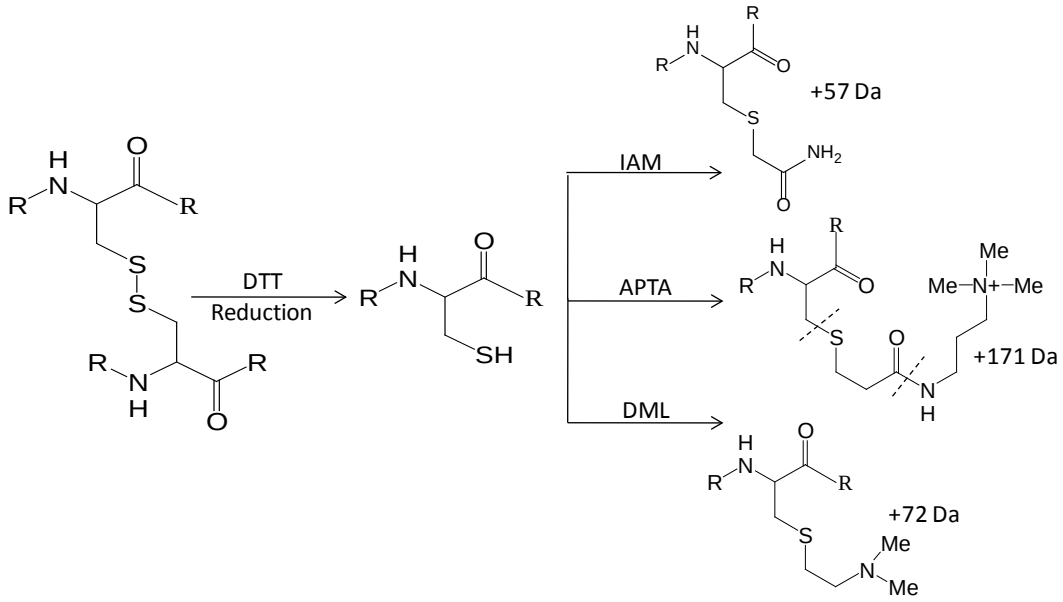

Figure 6.1 DTT reduction and alkylation of cysteines, dotted lines represent cleavage sites upon ETD to form APTA reporter ions.

A series of cysteine-containing model peptides were derivatized using the three alkylating reagents, IAM, DML, and APTA (**Figure 6.1**), and subsequently characterized by CID and ETD mass spectrometry. The APTA alkylation reaction is extremely selective for available cysteines through the alkylation of their thiol groups following the reduction of the disulfide bonds.[33,40] The reaction, as shown in **Figure 6.1**, results in the addition of 170 Da and a permanent positive charge at each thiol. It is well known that molecules with fixed charges often have higher ionization efficiencies and thus yield more intense ions, enhancing the detection of low abundance cysteine-containing peptides in complex mixtures.[29, 30, 36] The APTA derivatization efficiencies are estimated to be approximately 70% based on the ratio of the summed areas of all APTA-derivatized products relative to the summed areas of all APTA-derivatized products and unmodified species detected in the full ESI mass spectra (data not shown).

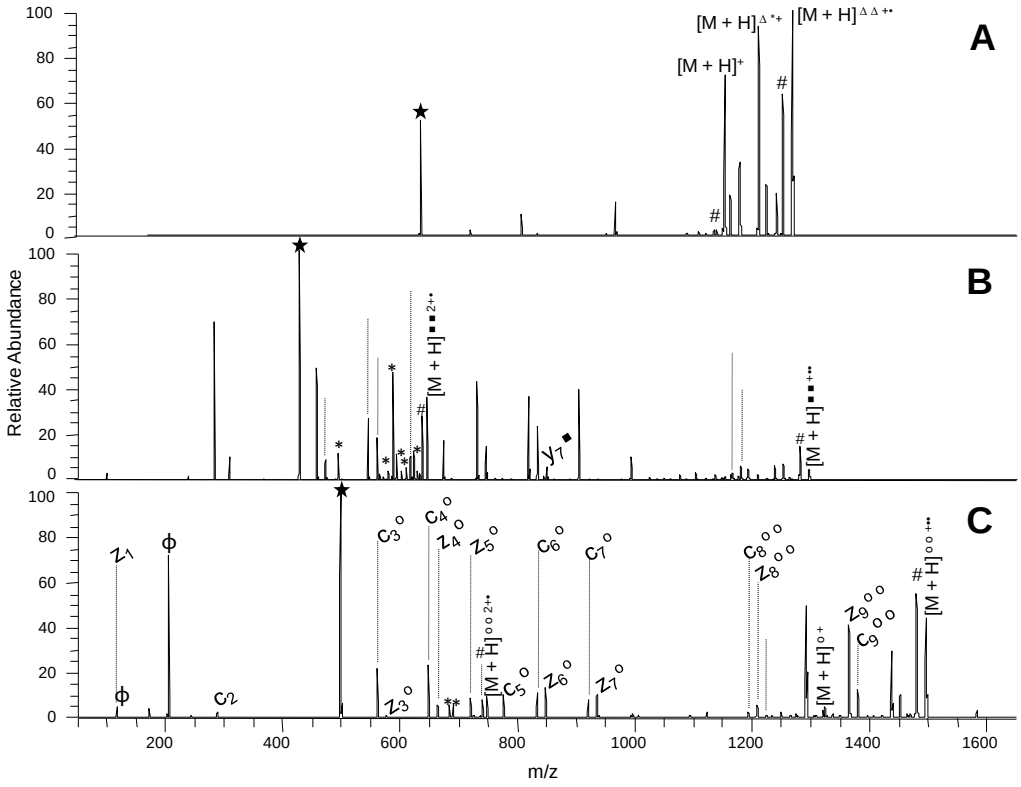

Figure 6.2 ETD product ion mass spectra of NRCSQGSCWN (A) doubly charged IAM-alkylated, (B) triply charged DML analog, and (C) triply charged APTA-derivatized. IAM-alkylation is denoted by Δ, DML-modification by ▪, APTA-derivatization by o, APTA-peptide reporter ions by ϕ, multiply charged fragment ions by * and neutral H_2O/NH_3 losses by #. The unmodified peptide is represented by M while the selected precursor ion by ★.

Representative examples of the ETD mass spectra obtained for a cysteine-containing peptide, NRCSQGSCWN, after each of the three derivatization procedures (IAM, DML, and APTA) are shown in **Figure 6.2**. The highest charge state of the modified peptides were selected for ETD, corresponding to 2+ for the IAM-derivatized peptide, which has neither a fixed charge site nor an additional highly basic site; 3+ for the DML-derivatized peptide, which has an additional basic site; and 3+ for the APTA-

derivatized peptide, which has a fixed charge site. As shown for the IAM-derivatized peptide, very few sequence ions (c_6, c_7, and c_8) are observed upon ETD, providing incomplete sequence coverage. In addition, these three *c*-ions allow the position of only one cysteine residue to be pinpointed. For the DML-modified peptide (**Figure 6.2B**), the ETD mass spectrum displays far more complementary *c* and *z* sequence ions, and the resulting sequence coverage (see inset of **Figure 6.2A**) is greatly improved although there are still gaps near the termini as the c_2, c_9, z_1, and z_8 ions were not observed. The ETD mass spectrum for the APTA-derivatized peptide reveals the entire *c* series as well as nearly the entire *z* series of product ions (**Figure 6.2C**). The latter mass spectrum provides the most extensive sequence coverage and thus affords confident assignment of the peptide sequence, along with identification of the positions of both cysteine residues. Moreover, upon ETD the formation of fragment ions in multiple charge states is minimized for the APTA-modified peptides as illustrated in **Figure 6.2C**. Upon DML derivatization, a significant number of multiply charged product ions crowd the 500-700 *m/z* range which is consistent with previous results for ETD of DML analogs of other cysteine-containing peptides.[28]

A comparison of the ETD and CID results for a series of model peptides derivatized by the three methods is summarized in **Figure 6.3**. The number of sequence ions (*c*-, *z*-, and *y*-type ions for ETD and *a*-, *b*-, and *y*-type ions for CID) are displayed in bar graph form for the highest charge states of each modified peptide. The ETD efficiencies and sequence coverages were significantly diminished for the lower charged peptide cations as the dominant reaction pathway was charge reduction without dissociation (data not shown). For the ETD results, the APTA-modified peptides

typically yielded the greatest number of *c*, *z*, and *y* ions. The sequence coverage of the APTA-modified peptides was consistently greater than that of the IAM-alkylated peptides and usually better than that obtained from the DML-modified peptides. For example, for bactenecin (RLCRIVVIRVCR), ETD of the triply charged IAM-alkylated peptide produced one *y* ion, eight *c* ions, and six *z* ions, ETD of the DML-derivatized peptide (5+) yielded one *y* ion, nine *c* ions, and nine *z* ions, and ETD of the APTA-modified peptide (5+) produced seven *y* ions, eleven *c* ions, and eleven *z* ions. Among the three types of derivatized peptides evaluated in the present study, the overall sequence coverage obtained upon ETD of the IAM-alkylated peptides was the least comprehensive as these derivatized peptides were predominantly observed in lower charge states due to the lack of additional highly basic or fixed charge sites. As previously noted, precursor cations in lower charge states tend to favor non-dissociative processes upon electron attachment.[8,12,13]

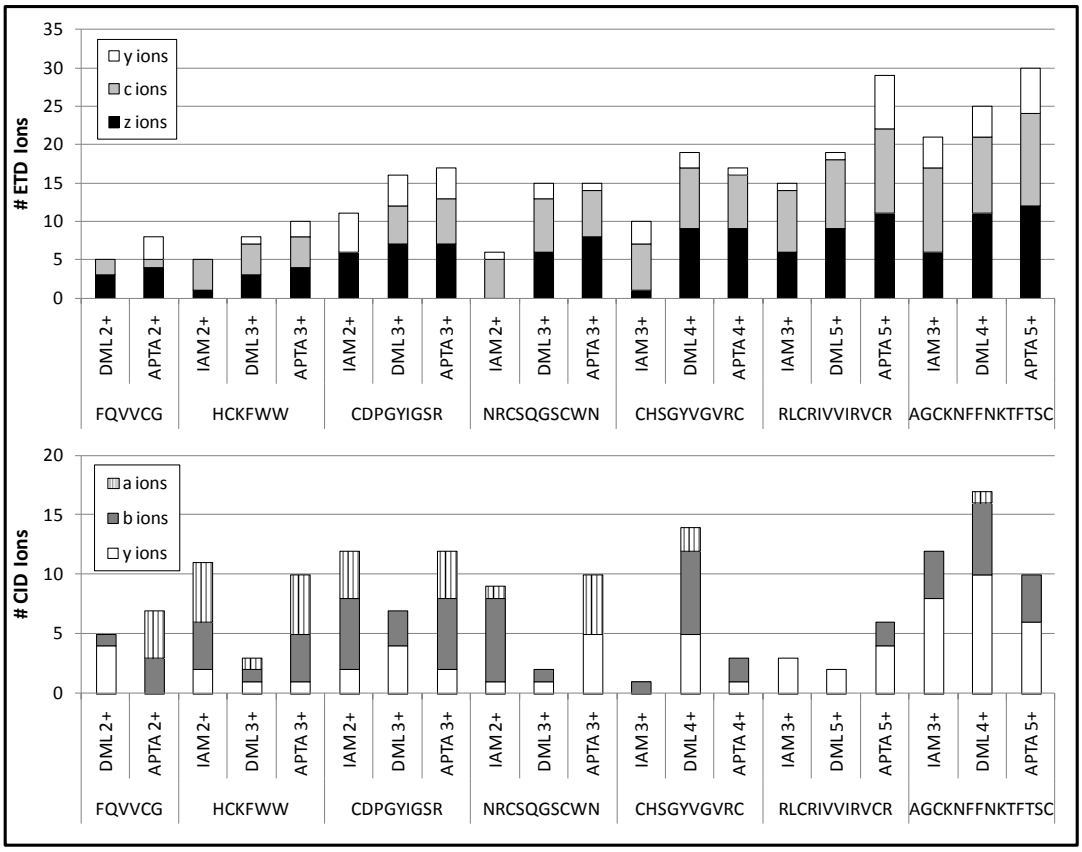

Figure 6.3 Number of diagnostic ions produced by ETD and CID for model peptides upon modification; IAM=IAM-alkylated, DML=dimethyl Lys analog, APTA=APTA-derivatized.

Two contributing factors led to the observed enhancement in sequence coverage for the APTA-peptides relative to the DML-peptides. First, the APTA-modified peptides frequently undergo backbone cleavages close to the peptide termini upon ETD that were not routinely observed for the DML-peptides. These include cleavages between the terminal and penultimate residues leading to the formation of the single C-/N-terminal residue ion (e.g., c_1, z_1 and y_1) and their complementary fragment ions (e.g., z_{n-1}, c_{n-1}). Second, the production of the lower energy *y* fragment ion series upon ETD is enhanced

for the APTA-peptides. The *y* fragment ions are commonly produced by CID, not ETD, but they are in fact reasonably abundant product ions observed upon ETD of the APTA-peptides. The increase in abundance of *y* ions for the APTA-peptides may be related to the presence of fixed charge sites at all cysteine residues. It has been proposed that the formation of *y*-type ions upon ECD is enhanced for peptides that can adopt zwitterionic structures as well as those peptides that have basic residues at the C-terminus or those with proline residues (i.e. peptides in which protons can migrate to proline amide or other backbone amide positions).[37] Peptides that can adopt zwitterionic structures promote greater proton mobility, thus allowing proton localization at the backbone amide nitrogen.[36] The highly basic sites introduced at cysteine residues encourage the formation of zwitterions, presumably increasing the production of the *y*-type ions. Moreover, as ETD does not frequently yield products from the N-terminal side of proline residues,[7] there are often gaps in the sequence coverage. This shortcoming can be compensated (albeit at the expense of more elaborate data collection) by undertaking complementary CID analysis of the peptides because CID often causes enhanced cleavages N-terminal to proline residues.[38-40] However, the production of *y* ions stemming from cleavages N-terminal to proline residues was common upon ETD of the APTA-peptides in the present study, thus providing another set of N-terminal ions to improve the confidence in identifying the sequences. Both of these improvements can be visualized in Figure 2, in which the number of *y* ions produced is increased for the APTA-peptides in comparison to the DML-peptides and IAM-peptides.

Another way of assessing the analytical value of the ETD results for the IAM-, DML- and APTA-derivatized peptides entails the comparison of percent fragmentation

values, defined by Coon et al. as the number of observed *c*- and *z*-type ions for a sequence divided by the total number of theoretically possible *c*- and *z*-type fragment ions.[41] Larger percent fragmentation values signify higher sequence coverage. The percent fragmentation is plotted as a function of the *m/z* values of the peptides in **Figure 6.4**. ETD of the APTA-peptides generally affords higher percent fragmentation values than that obtained for the IAM-peptides or DML-peptides. The decrease in percent fragmentation with increasing precursor *m/z* has been previously documented and leads to a precursor *m/z* limit for successful ETD.[41]

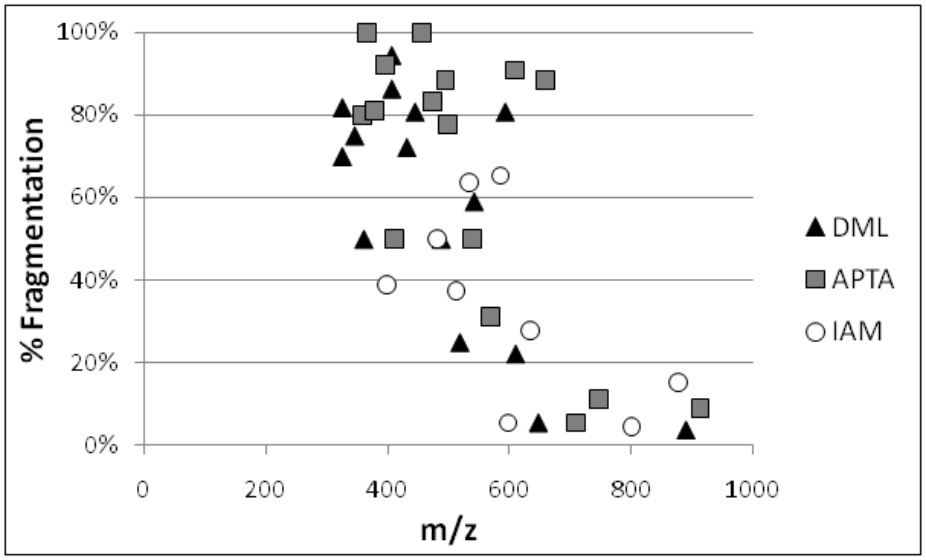

Figure 6.4 Percent fragmentation obtained upon ETD of model peptides as a function of precursor *m/z* and derivatization method. Percent fragmentation is defined as the number of observed fragment ions divided by the total number of possible fragment ions (all *c*- and *z*-type ions).

The ETD dissociation efficiencies, defined as the sum of the abundances of the *z*, *c*, and *y* ions divided by the total ion abundance (abundance of all remaining ions

including any backbone fragments, product ions due to neutral losses, reporter ions, charge reduced species and any surviving precursor ion), were calculated and expressed as percentages for the highest charge state (as listed in **Figure 6.3**) for the APTA-, DML-, and IAM-derivatized peptides. The dissociation efficiencies for the APTA- and DML-peptides were comparable, averaging 47% and 53% for the series of model peptides, respectively, and only 18% for the IAM-peptides.

In addition to the benefits of the greater sequence coverage afforded upon ETD of the APTA-peptides, a group of characteristic reporter ions is also produced. Two reporter ions of *m/z* 117 and 205 offer a secondary means of readily pinpointing cysteine-containing peptides. These two reporter ions are likely produced through cleavage of the APTA moiety (*m/z* 117 and 205), as illustrated in **Figure 6.1** by the dashed cleavage lines. These ions are produced in reasonably high abundance for all charge states of the APTA-peptides, thus allowing consistent tracking of all of the APTA-peptides as seen in **Figures 6.2** and **6.5**.

Based on the similar or better sequence coverage obtained for the peptides after APTA derivatization, the attachment of fixed charges (e.g., via APTA) rather than basic sites (e.g., via DML) was pursued for the remainder of the comparative study.

6.4.2 ETD of BSA Tryptic Digest

Following the assessment of the results obtained for the series of model peptides derivatized by the three reagents, tryptic peptides of bovine serum albumin were similarly analyzed. The BSA tryptic peptides were either alkylated with IAM or converted to APTA analogs for analysis by ETD and CID mass spectrometry. As anticipated, the IAM

alkylation procedure, although efficient, led to the production of derivatized peptides in lower charge states (predominantly 2+ or 3+ depending on the length of the peptide) upon ESI than did the APTA derivatization (charge states of 3+ to 7+ observed). Consequently, ETD of the IAM-alkylated peptides largely resulted in non-dissociative charge transfer yielding singly charged radical peptides and formation of few sequence ions. An example can be seen in **Figure 6.5A** for the tryptic peptide YNGVFQECCQAEDK, in which only seven *c*- and *z*-type ions and two *y*-type ions are observed, thus diminishing the identification of the peptide. In comparison, ETD of the analogous APTA-modified peptide led to the production of far more sequence ions, including complete series of *c*, *z* ions, and *y* ions and minor abundances of charge-reduced peptides (**Figure 6.5B**).

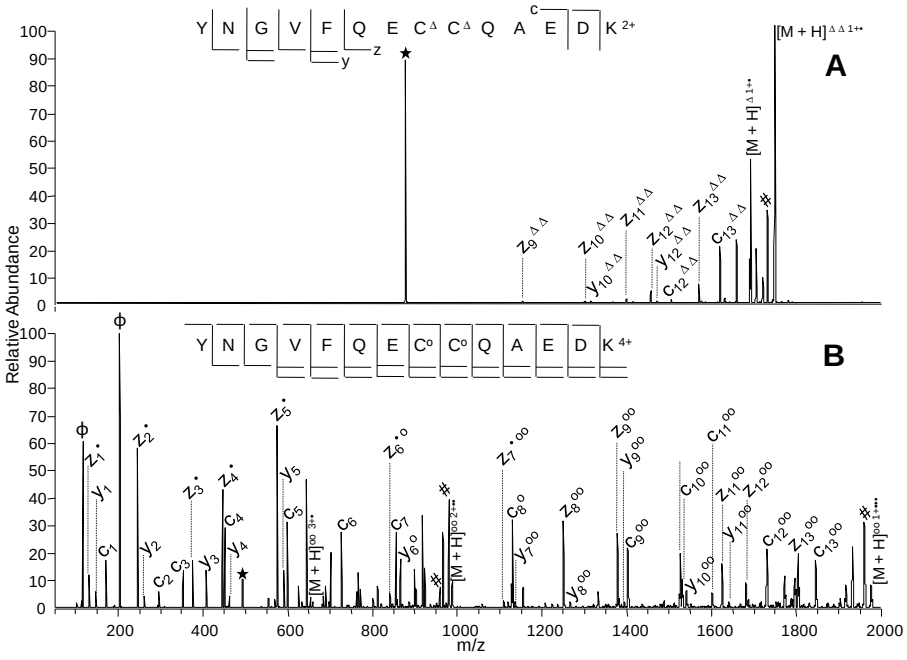

Figure 6.5 ETD product ion mass spectra of the BSA tryptic peptide YNGVFQECCQAEDK for the (**A**) doubly charged IAM-alkylated, and the (**B**) quadruply charged APTA-derivatized. IAM-alkylation is denoted by Δ, APTA-derivatization by o, APTA-peptide reporter ions by ϕ, and neutral H_2O/NH_3 losses by #. The unmodified peptide is represented by M, while the precursor ion by ★.

The results for the entire tryptic digest of BSA are summarized in **Figure 6.6** which shows the number of diagnostic *z*, *c*, and all *y* and *b* ions upon ETD and *y*, *b* and *a* ions upon CID for twenty-five IAM- and APTA-derivatized tryptic peptides. In order to maximize the number of diagnostic ions formed, ETD and CID were undertaken on the highest charge states observed for each of the IAM- and APTA-peptides. The APTA derivatization allowed analysis of many peptides that were either below the mass range or too low in abundance to be isolated upon conventional alkylation. For example, the two smallest peptides, CASIQK and GACLLPLK, were solely identified following the APTA derivatization/ETD strategy because without the APTA addition the *m/z* values of the multiply charged states of these small peptides fell below the *m/z* range analyzed. Moreover, the singly charged analogs with correspondingly higher *m/z* values do not produce charged products upon ETD. In addition, three peptides SLHTLFGELCK, SQYLQQCPFDEHVK, and AFSLQYLQQCPFDEHVKLVNELTEFAK, were not detected after the IAM-alkylation procedure. In contrast, upon APTA derivatization these three peptides exhibited greater ionization efficiencies that allowed their detection and analysis in the data-dependent LC run.

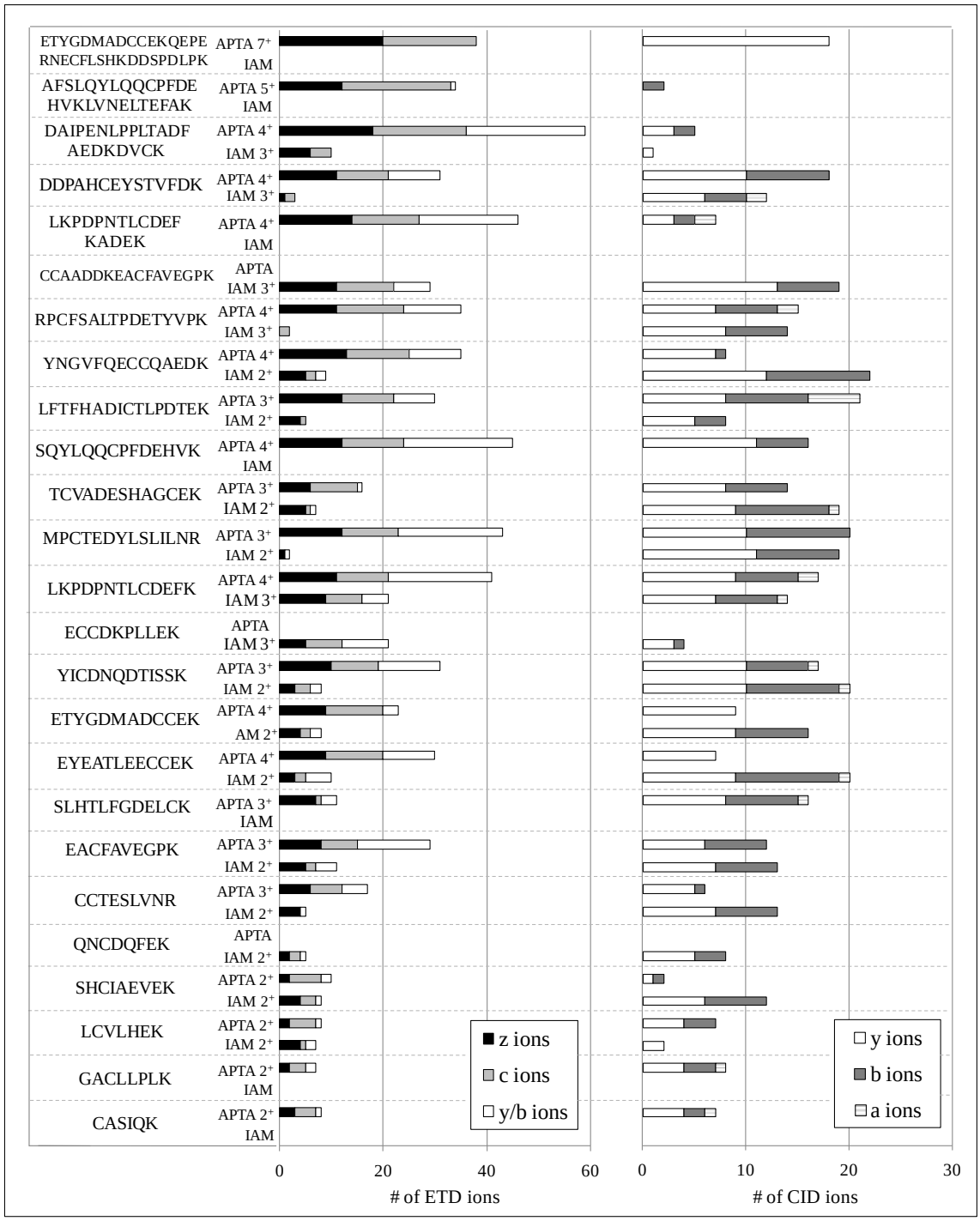

Figure 6.6 Number of diagnostic ions produced by ETD and CID for BSA tryptic peptides; IAM=IAM-alkylated, APTA=APTA-derivatized. If no bar is

present, that derivatized peptide was not detected in any charge state and could not be analyzed.

There were three cases where the APTA-derivatized tryptic peptides were not detected (QNCDQFEK, ECCDKPLLEK, and CCAADDKEACFAVEGPK). Upon derivatization of all available cysteines and a corresponding increase in the charge states of these peptides, the *m/z* values of the resulting derivatized peptides fall below *m/z* 400 (the lowest mass scanned in the full mass spectra to eliminate LC solvent peak interferences) and thus these ions would not be detected. Despite this shortcoming, ~50% of BSA was characterized through the investigation of solely the APTA-derivatized cysteine-containing peptides. For all cases in which both IAM- and APTA-derivatized peptides were observed, ETD of the fixed charge APTA-peptides provided significantly greater sequence coverage than that obtained for the IAM-derivatized peptides. The same trend does not hold true for the analogous CID results. In some cases, the IAM-alkylated peptides yielded better sequence coverage, whereas in other cases the CID of the APTA-peptides gave greater sequence coverage. These CID results were not unexpected considering that the APTA-derivatization procedure adds fixed charges, not mobile protons that facilitate many of the fragmentation pathways observed by CID.[42, 43]

As described earlier, a second way of assessing the analytical value of the ETD results for the derivatized peptides involves comparing the percent fragmentation values. The percent fragmentation values are plotted as a function of the *m/z* values of the derivatized tryptic peptides in **Figure 6.7**. ETD of the APTA-peptides generally affords higher percent fragmentation values than that obtained for the IAM-peptides, again reflecting the merits of employing the fixed charge derivatization procedure for cysteine-containing tryptic peptides.

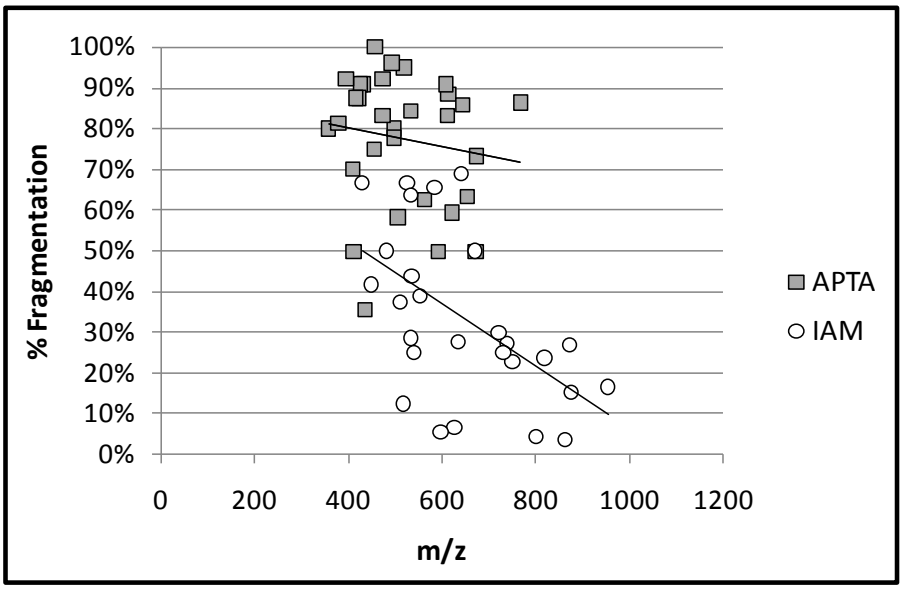

Figure 6.7 Percent fragmentation obtained upon ETD of tryptic peptides from BSA as a function of precursor *m/z* and derivatization method. Percent fragmentation is defined as the number of observed fragment ions divided by the total number of possible fragment ions (all *c*- and *z*-type ions).

The ETD dissociation efficiencies were also calculated for the series of derivatized tryptic peptides shown in **Figures 6.6** and **6.7**. The dissociation efficiencies for the APTA-peptides averaged 56%, whereas the dissociation efficiencies for the IAM-peptides averaged only 17%. As expected from the well-recognized charge state dependence of ETD, the more highly charged APTA-peptides yield greater relative abundances of informative sequence ions upon ETD. This consistent increase in sequence ions also improves the application of search algorithms used to identify peptide sequences, thus exploiting the advantage of fixed charge derivatization in conjunction with ETD. As proof of concept, the HPLC data obtained for the tryptic digests of both alkylations was submitted to SEQUEST search software using the bovine taxonomy

database. In both cases BSA was identified using the search parameters described in the experimental section with the highest probability score. For the IAM alkylation method, a total of 21 peptides were correctly identified with high confidence leading to a SEQUEST score of 582 and sequence coverage of 55%. For the APTA-modified BSA run, the search was limited to only APTA-modified cysteine peptides, yielding the correct identification of 15 peptides with a score of 3700 and 23% protein coverage. This 6-fold increase in score accompanied by a decrease in sequence coverage proves that the selective study of tagged Cys peptides from a complex mixture is a powerful identification database.

6.5 CONCLUSIONS

The cysteine-selective fixed charge reagent, APTA, offers advantages over other charge enhancement strategies and traditional alkylation procedures in conjunction with ETD mass spectrometric analysis of peptides. The APTA derivatization procedures yields peptides in higher charge states upon ESI, and ETD of the APTA-modified peptides results in the formation of nearly complete series of *c*, *z*, and *y* ions. The APTA modification increased the sequence coverage by enhancing fragmentation near the peptide termini as well as production of *y* ions not typically observed upon ETD. ETD of the APTA-peptides also resulted in higher fragmentation percentages (number of observed fragment ions/number of theoretically possible fragment ions) and greater dissociation efficiencies than obtained for conventional IAM-alkylated peptides. Generation of a pair of reporter ions upon ETD of the APTA-peptides offered a convenient means of tracking the cysteine-containing peptides.

6.6 REFERENCES

(1) McLafferty, F. W.; Venkataraghavan, R.; Irving, P. *Biochem. Biophys. Res. Commun.* **1970**, *39*, 274-278.

(2) Wells, J. M.; McLuckey, S. A. *Methods Enzymol.* **2005**, *402*, 148-185.

(3) Papayannopoulos, I. A. *Mass Spectrom. Rev.* **1995**, *14*, 49-73.

(4) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 190-208.

(5) Brodbelt, J. S.; Wilson, J. J. *Mass Spectrom. Rev.* **2009**, *28*, 390-424.

(6) Reilly, J. P. *Mass Spectrom. Rev.* **2009**, *28*, 425-447.

(7) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.

(8) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563-573.

(9) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528-9533.

(10) Srikanth, R.; Wilson, J.; Bridgewater, J. D.; Numbers, J. R.; Lim, J.; Olbris, M. R.; Kettani, A.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1499-1506.

(11) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Anal. Chem.* **2007**, *79*, 477-485.

(12) Ben Hamidane, H.; Chiappe, D.; Hartmer, R.; Vorobyev, A.; Moniatte, M.; Tsybin, Y. O. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 567-575.

(13) Ben Hamidane, H.; He, H.; Tsybin, O. Y.; Emmett, M. R.; Hendrickson, C. L.; Marshall, A. G.; Tsybin, Y. O. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1182-1192.

(14) Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *Anal. Chem.* **2005**, *77*, 5662-5669.

(15) Iavarone, A. T.; Paech, K.; Williams, E. R. *Anal. Chem.* **2004**, *76*, 2231-2238.

(16) Chamot-Rooke, J.; Malosse, C.; Frison, G.; Turecek, F. *J. Am. Soc. Mass. Spectrom.* **2007**, *18*, 2146-2161.

(17) Chamot-Rooke, J.; van der Rest, G.; Dalleu, A.; Bay, S.; Lemoine, J. *J. Am. Soc. Mass. Spectrom.* **2007**, *18*, 1405-1413.

(18) Gunawardena, H. P.; Gorenstein, L.; Erickson, D. E.; Xia, Y.; McLuckey, S. A. *Int. J. Mass spectrom.* **2007**, *265*, 130-138.

(19) Li, X.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1514-1526.

(20) Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W. *J. Am. Chem. Soc.* **2001**, *123*, 9792-9799.

(21) Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 6407-6420.

(22) Horn, D. M.; Ge, Y.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 4778-4784.

(23) Oh, H.; McLafferty, F. W. *Bull. Korean Chem. Soc.* **2006**, *27*, 389-394.

(24) Xia, Y.; Han, H.; McLuckey, S. A. *Anal. Chem.* **2008**, *80*, 1111-1117.

(25) Han, H.; Xia, Y.; McLuckey, S. A. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1567-1573.

(26) Ledvina, A. R.; Beauchene, N. A.; McAlister, G. C.; Syka, J. E. P.; Schwartz, J. C.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2010**, *82*, 10068-10074.

(27) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. *Anal. Chem.* **2007**, *79*, 9243-9252.

(28) Ueberheide, B. M.; Fenyo, D.; Alewood, P. F.; Chait, B. T. *Proc. Nat. Acad. Sci. U.S.A.* **2009**, 1-6.

(29) Ren, D.; Julka, S.; Inerowicz, H. D.; Regnier, F. E. *Anal. Chem.* **2004**, *76*, 4522-4530.

(30) Roberts, K. D.; Reid, G. E. *J. Mass Spectrom.* **2007**, *42*, 187-198.

(31) Huang, Z.-H.; Wu, J.; Roth, K. D. W.; Yang, Y.; Gage, D. A.; Watson, J. T. *Anal. Chem.* **1997**, *69*, 137-144.

(32) Wang, D.; Kalb, S. R.; Cotter, R. J. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 96-102.

(33) Sadagopan, N.; Watson, J. T. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 107-119.

(34) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 377-384.

(35) Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2009**, *81*, 3645-3653.

(36) Li, J.; Ma, H.; Wang, X.; Xiong, S.; Dong, S.; Wang, S. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2608-2612.

(37) Tsybin, Y. O.; Haselmann, K. F.; Emmett, M. R.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1704-1711.

(38) Vaisar, T.; Urban, J. *J. Mass Spectrom.* **1996**, *31*, 1185-1187.

(39) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Science* **1990**, *248*, 201-204.

(40) Barinaga, C. J.; Edmonds, C. G.; Udseth, H. R.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 160-164.

(41) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. *Molecular and Cellular Proteomics* **2007**, *6*, 1942-1951.

(42) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365-8374.

(43) Sadagopan, N.; Watson, J. T. *J. Am. Soc. Mass. Spectrom.* **2001**, *12*, 399-409.

Chapter 7

Introduction of a Hydrazone Bond to Facilitate Mapping Novel Protein Surface Accessibility through Electron Transfer Dissocation

7.1 OVERVIEW

A protein's surface is especially important for its role in protein-protein interactions and protein-ligand binding. Mass spectrometry can be used to give low resolution structural information when used in combination with derivatization methods that target surface accessible amino acid residues. However, pinpointing the resulting modified peptides upon enzymatic digestion of the surface-modified protein is challenging due to the complexity of the peptide mixture and low abundance of modified peptides. Here a novel reagent, NN, is presented that allows facile identification of all modified surface residues through a preferential cleavage upon activation by electron transfer dissociation coupled with a data-dependent MS^3 scan to pinpoint the modified residue in the peptide sequence. Using this approach, the standard protein, ubiquitin, is used to show the correlation between percent reactivity and solvent accessibility. In addition, the surface accessibilities of two other proteins, eIF4E and PARP-1 Domain C, were studied.

7.2 INTRODUCTION

In the past decade the number of protein sequences without solved structures in the Protein Data Bank library has increased dramatically.[1] Determination of the correlation between protein structure and function remains a primary objective of biological research, thus motivating the development of advanced analytical tools for

unraveling the three dimensional structures of proteins. The most common techniques used to determine higher order structure of a protein are nuclear magnetic resonance (NMR) and X-ray crystallography; however these techniques are not universal due to practical limitations related to protein size, the inability to crystallize certain proteins, or limited sample amounts. Due to these restrictions, the development of mass spectrometry-based strategies has gained traction due to their speed and sensitivity and have shown great promise for low resolution protein structure analysis.

A number of protein labeling techniques have been used in combination with tandem MS analysis to provide low resolution structural information including hydrogen-deuterium exchange (HDX), crosslinking, and covalent chemical modifications prior to proteolytic digestion.[2] HDX provides the most detailed information about a structure as it probes the entire protein backbone; however, spatial resolution can be greatly limited due to back-exchange of the deuterium to hydrogen before analysis[3] or hydrogen/deuterium scrambling during tandem MS.[4] Electron-based dissociation methods have been able to overcome much of the error due to scrambling, but the limitations due to back-exchange still remain great.[5, 6] Chemical modifications are used to covalently label specific or nonspecific amino acid side chains, thus eliminating the problem of back-exchange and scrambling.[2] The foundation of this method is based on the premise that amino acids that are exposed to solvent and are therefore accessible to a chemical labeling agent will be modified, whereas those that are buried will be modified slowly or not at all. This type of labeling provides information about the identities of the amino acids accessible to the reagent in solution, resulting in a low resolution map of the protein's surface. In particular, the protein's surface and accessibility of amino acid side

chains reflects their potential participation in protein-protein or protein-ligand interactions. The chemical modification methods are frequently combined with a bottom-up approach in which the modified proteins are subsequently enzymatically digested to facilitate MS/MS analysis. Crosslinking, a specific form of chemical modification, provides information about structure by creating new intramolecular or intermolecular bonds between specific amino acids with distance constraints on the location of the two amino acids linked due to the size of the linker.[7] Since new bonds are formed between contact areas of the protein, even if remote in terms of the primary sequence, sequencing these sites through tandem MS can prove challenging. Identification of the crosslinked peptides provides distance constraints that can be used to re-construct the contact points and conformation of the protein(s). With single covalent modifications of amino acids (not formation of crosslinks), the analysis is often more straightforward as the modification can be treated as a post-translational modification, thus allowing most searching algorithms to be used for the identification of the digested modified peptides. In this case, the identification of the modified residues reveals the exposed regions of the protein relative to the inaccessible regions, therefore reflecting conformational information.

Lysine is one of the most targeted amino acids for chemical labeling due to its intrinsically high reactivity, thus making it amenable to efficient modification. Moreover, the positively charged, polar side chain of lysine under physiological pH conditions means that lysines are more often located on the hydrophilic surface of proteins and consequently more often involved in protein-protein or protein-ligand interactions.[2] Protein-protein and protein-ligand interactions can be studied through this

covalent modification strategy based on monitoring the differential reactivity of selected residues in the presence/absence of the interacting protein or ligand, indicating their involvement in a binding interaction. Several different Lys-specific chemical modification methods have been reported such as aminoacetylation[8-13], amidination,[14, 15] and many biotin labeled reagents.[16-20] These modifications have been used to characterize the topology of multiple proteins, protein-ligand complexes and protein-protein interactions.

The development and application of new crosslinking and surface accessibility strategies has been impeded by the difficulty of detecting the low abundance crosslinked or modified peptides amidst a large array of more abundant unmodified peptides produced upon enzymatic digestion of the proteins or protein complexes. This detection problem has been addressed by efforts to selectively enrich the modified peptides,[21-27] incorporation of isotopic labels in the reagents to give the modified peptides distinctive isotopic signatures,[28-32] and development of selectively cleavable reagents that yield a traceable neutral loss or reporter ions upon MS/MS analysis.[33-37] In the context of surface accessibility studies, Reid et al. has done this using a 'fixed charge' sulfonium ion with specificity for methionine[38], cysteine[39] and lysine[40] targets where upon collision induced dissociation (CID) the exclusive loss of a dialkylsulfide (e.g. a neutral loss of 62 Da) is indicative of a modified species. Our group has previously found that a N-N hydrazone bond was selectively cleaved upon electron transfer dissociation.[37] Bis-arylhydrazone crosslinked peptides were created by reacting succinimidyl 4-formylbenzoate modified peptides with succinimidyl 4-hydrazinonicotinate acetone hydrazone modified peptides. These crosslinked peptides were then quickly identified

through the production of two partially modified peptides, one even-electron product ion and one odd-electron product ion, following cleavage of the N-N hydrazone bond. Here we describe a new reagent, NN (see **Figure 7.1**), that modified Lys and N-terminal residues by incorporation of a N-N hydrazone bond. Upon electron transfer dissociation (ETD), the preferential cleavage of the N-N bond leads to a dominant loss of 93 Da from all charge-reduced species that can readily distinguish all modified peptides from unmodified ones. The characteristic loss of 93 Da can be used to trigger a subsequent CID event (MS^3) to sequence the modified peptides and identify their locations within the protein sequence. Because all modified peptides are uniquely pinpointed based on the characteristic neutral loss from the charge reduced species and this product retains a unique mass tag too yet in a lower charge state, the resulting MS/MS/MS fragment ions are produced in lower charge states that increase the confidence of their identification. To initially test the method and reagent, a model protein (ubiquitin) was used to evaluate the feasibility of the surface interrogation. Following the proof of concept, the surface accessibilities of two biologically relevant proteins, wheat eIF4E and PARP 1 Domain C, were mapped.

7.3 EXPERIMENTAL

7.3.1 Materials and Reagents

Ubiquitin, proteomics grade trypsin, endoproteinase Arg C and Glu C were purchased from Sigma (St. Louis, MO). PARP 1 Domain C protein was provided by Dr. Hung-wen Liu (Department of Pharmacy, University of Texas at Austin) and eIF4E was provided by Dr. Karen Browning (Department of Chemistry and Biochemistry,

University of Texas at Austin). All other chemicals and solvents were purchased from Fisher scientific (Fairlawn, NJ).

7.3.2 Synthesis of (E)-2,5-dioxo-1-((4-((2-(pyridine-2-yl)hydrazono)methyl)benzoyl)oxypyrrolidine-3-sulfonic acid (NN)

The surface accessibility reagent, NN, was synthesized in house as summarized in the **Figure 7.1** below. NN was synthesized by refluxing 7.3 mmol of 2-hydrazinopyridine and 9.7 mmol of 4-formylbenzoic acid in 20 mL of toluene for four hours. A yellow precipitate was isolate by vacuum filtration and washed with toluene and ethyl acetate (82.0%). One equivalent of the isolated hydrazonal intermediate was coupled with two equivalents of N-hydroxysulfosuccinimide using two equivalents of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide and a catalytic amount of 4-dimethylaminopyridine in dimethylformamide for 24 hours. Upon completion, the reaction was rotovaped and washed with dichloromethane and water. The product was triturated in hexanes before isolating a dark yellow solid (46.8%).

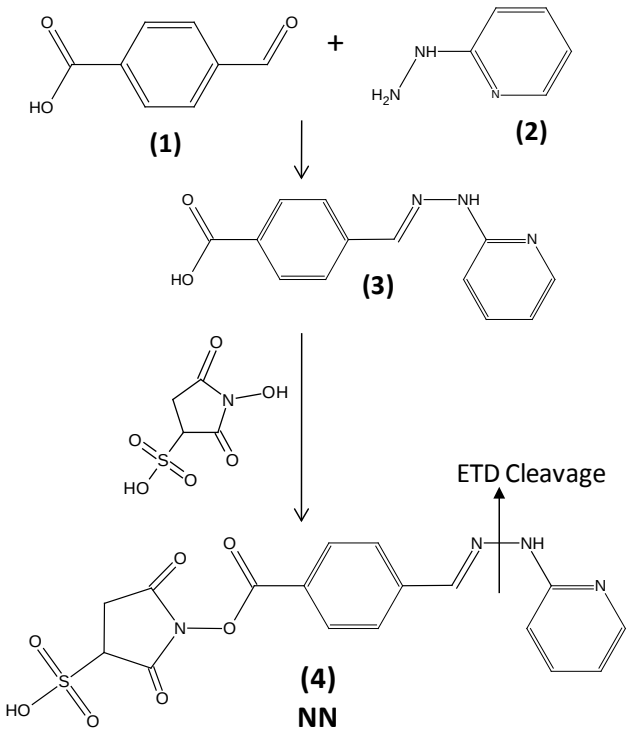

Figure 7.1 Synthesis of NN; NN was synthesized by coupling 4-formylbenzoic acid **(1)** and 2-hydrazinopyridine **(2)** and via hydrazone bond formation. The hydrazone intermediate **(3)** was subsequently reacted with sulfo-N-hydroxysuccinimide to install the amine-reactive NHS terminus producing NN **(4)**; preferential ETD cleavage is marked indicating the loss of 93 Da from all charge reduced species.

7.3.3 Derivatization and Sample Preparation

Each protein (1 mM, 25µL) was mixed with NN at a variety of molar ratios relative to NN (20 mM) in PBS buffer at pH 7.2 - 7.4. The protein:NN ratios were varied from 1:2 to 1:10. All reactions were carried out at room temperature overnight and cleaned up using 10 kDa MWCO filters the following morning. Derivatized samples were then split into three aliquots for digestion. For tryptic digestion, the derivatized protein (8 nmol) was diluted with 100 mM NH_4HCO_3 and digested with 1 mg/mL trypsin in 1 mM HCl in a 1:100 w/w ratio of

protein to trypsin overnight at 37°C. For GluC digestion, the derivatized protein (8 nmol) was diluted with the GluC Reaction Buffer from BioLabs, consisting of 50 mM Tris-HCl and 0.5 mM Glu-Glu buffer at pH 8.0, and digested with 1 mM GluC in water in a 1:20 w/w ratio overnight at 37°C. For ArgC digestion, the derivatized protein (8 nmol) was diluted in 50 mM NH_4HCO_3 at a pH of 8.0 with a small addition of 20 mM $CaCH_3COOH$ to enhance digestion, and then digested using a 1:16 w/w molar ratio of protein to ArgC at 37°C overnight. The digested samples were diluted to 10 μM before ESI analysis with 49.5/49.5/1 H_2O/MeOH/Acetic Acid.

7.3.4 Mass Spectrometry and Liquid Chromatography

All experiments were undertaken on a ThermoFisher LTQ XL linear ion trap mass spectrometer (San Jose, CA) equipped with an ETD unit. Direct infusion analysis on the LTQ XL was done using an online nanoESI setup as previously described using a flow rate of 3 μL/min at a concentration of 10 μM in 49.5:49.5:1 MeOH:H_2O:Acetic acid.[41] An ESI voltage of 2 kV and a heated capillary temperature of 180^o C were used for experiments on the LTQ XL. Liquid chromatography was performed using a RLSC Dionex UltiMate 3000 system (Sunnyvale, CA). An Agilent ZORBAX 300Extend-C_{18} column (Santa Clara, CA) (150 × 0.3 mm, 3.5 μm particle size) was used for all separations. Eluent A consisted of 0.1% formic acid in water and eluent B 0.1% formic acid in acetonitrile. A linear gradient from 5% eluent B to 40% eluent B over 65 min at 0.3 μL/min was used. Injections of approximately one picomole were used for each digested sample. For all LC-MS/MS runs, the first event was the full mass scan (*m/z* range of 400 – 2000) followed by five sets of consecutive MS/MS events on the five most abundant ions from the full mass scan. The first event in each set was acquisition of an ETD spectrum using an electron transfer reaction time of 100 ms. Following ETD, the most abundant fragment ion was subsequently activated by CID using a *q*-value of 0.25, a normalized collisional energy of 35%, and a collision activation of 30 ms. The maximum injection time for all events was set to 100 ms, and each mass spectrum and tandem mass spectrum was the average of five microscans.

7.3.5 Determination of Surface Accessibility

Upon identification as a modified peptide from the ETD spectrum based on the characteristic loss of 93 Da, peptides were sequenced manually using the subsequent CID fragmentation pattern of the charge-reduced species. The percent reactivity was calculated based on the sum of the peak areas of all peptides containing a modified residue based on the total ion chromatographic (TIC) profiles and integrated using QualBrowser, divided by the sum of the area of all peptides containing the unmodified and modified residue as shown in equation 7.1.

$$Equation\ 7.1: \%\ Reactivity = \frac{\sum Area\ of\ all\ peptides\ containing\ modified\ residue\ n}{\sum Area\ of\ all\ peptides\ containing\ residue\ n}$$

For proteins where a known structure was not available, ITASSER was used to predict a structure based on the primary sequence of the protein.[1] Theoretical surface accessibilities for all Lys side chains and the N-terminus were calculated using GetArea software online[42] with all parameters set as the default values based on known or ITASSER model structures as indicated.

7.4 RESULTS AND DISCUSSION

7.4.1 Design of NN Surface Accessibility Reagent

NN was designed to selectively react with primary amines, such as the ε-amino groups on lysine side chains and the N-terminus via conventional N-hydroxysuccinimide

(NHS) coupling. Second, NN contains a N-N hydrazone bond that has previously been shown to preferentially cleave upon ETD, thus facilitating the identification of modified peptides based on an easily monitored MS/MS reaction. For NN, the characteristic fragmentation upon ETD results in the neutral loss of 93 Da from the charge-reduced precursor and the number of consecutive neutral losses is indicative of the number of modified residues. This pathway is the dominant fragmentation pathway of NN-modified peptides upon ETD, and it allows the convenient implementation of a data dependent scan mode in which the acquisition of a CID spectrum can be triggered for the most abundant fragment ion produced upon ETD. The resulting CID (MS^3) spectrum allows the peptide to be sequenced and the location of the modification to be pinpointed. For the reactions of NN with each protein, the protein:NN ratios were varied over a range of values in order to optimize reaction efficiencies while maintaining tertiary protein structure. At lower molar ratios, only the most accessible sites are modified and the structural integrity is most readily maintained. With higher molar ratios, reaction efficiencies at less accessible modification sites increased, thus enhancing detection of those sites.

7.4.2 Ubiquitin

A model protein, ubiquitin, was utilized to test the efficiency of the NN reaction and its suitability for measurement of surface accessibility. Ubiquitin contains eight primary amine sites including seven Lys and the N-terminus. The reaction of ubiquitin with the NN reagent is efficient, as evidenced by the ESI-mass spectrum of the protein after modification (see **Figure 7.2**).

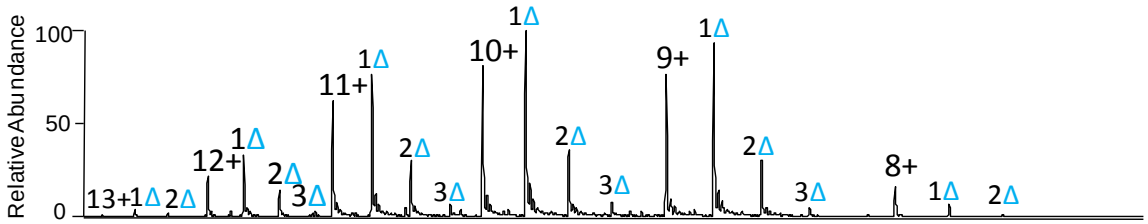

Figure 7.2 ESI mass spectrum of ubiquitin after reation with NN at a 1:2 protein: NN molar ratio. Triangles indicate NN modification.

Subsequent tryptic digestion and LC-MS/MS analysis results in identification of 9 unmodified peptides plus 14 modified peptides, the latter based on the characteristic N-N bond cleavage upon ETD followed by data-dependent CID. The abundances of the unmodified peptides and NN modified counterparts are quantified through manual integration of peak areas in the TIC profile using QualBrowser software, and the results are summarized in **Table 7.1** along with the solvent accessibilities estimated from the GetArea algorithm.

Residue	GetArea SA %	% Reactivity		
	UBI1	Trypsin	GluC	ArgC
Met 1	14	26±13	25±22	34±23
Lys 6	59	96±6	71±10	56±30
Lys 11	63	22±9	13±12	14±27
Lys 27	11	1±0.5	0	0
Lys 29	36	15±7	0	0
Lys 33	47	59±30	71±29	0
Lys 48	56	58±34	66±29	35±23
Lys 63	77	48±8	81±2	59±13

Table 7.1 Comparison of Surface Accessibility to Percent Reactivity for Ubiquitin at a 1:2 protein:reagent molar ratio. Solvent Accessibility was calculated using GetArea for pdb UBI1. % Reactivity was calculated using equation 7.1.

When the solvent accessibilities, according to GetArea based on the previously determined tertiary structure, of these sites are compared, they range from 11 to 77%. Ubiquitin was reacted with NN using four protein/NN molar ratios from 1:2 to 1:10, yielding up to three modifications for the low molar ratio of 1:2 and up to five modifications for the higher molar ratios of reagent to intact protein. The higher protein/NN ratios were used to enhance the modification and subsequent detection of some of the less accessible sites that might otherwise be non-detectable under more limiting protein:NN reaction conditions. These results are in general agreement with the expected solvent accessibilities of the amine sites given that four Lys residues have solvent accessibilities greater than 50% and one Lys is slightly lower at 47% based on the GetArea predictions. These five Lys residues are situated on the surface of the protein and therefore are more accessible to modification whereas the other three primary amines reside on the interior of the tertiary structure, making their side-chains largely inaccessible and unreactive. Three proteases, trypsin, GluC and ArgC, were used to digest the protein, thus facilitating identification of the modified peptides in a bottom-up approach. Although trypsin is the most popular protease for conventional proteomics strategies, it proves to be problematic for proteins modified at lysine sites, such as the NN-modified proteins in the present study, in which the modification disrupts proteolytic cleavage after the lysines. However, trypsin does give the most comprehensive sequence

coverage of all three digests and thus provided information about some of the more inaccessible residues.

Upon covalent modification of any protein, slight changes in structure are possible following derivatization. Thus to preserve the structure of the protein as much as possible, lower protein/NN molar ratios are used to limit the number of modifications per protein. For ubiquitin, a 1:2 protein/NN ratio produced predominantly singly modified proteins with minor contributions of doubly and triply modified proteins, as shown in **Figure 7.2**. The solvent accessibilities of each of the eight possible modification sites of ubiquitin are compared to the percent modifications obtained from the NN reactions in **Table 7.1**.

Upon digestion, the modified peptides were easily distinguished from unmodified peptides using ETD through the preferential cleavage of the NN bond. This cleavage leads to a loss of 93 from all charge reduced species. In addition, sequential losses of 93 Da can be used to identify the number of modifications. For example, the N-terminal peptide for ubiquitin, MQIFVKTLTGK, has two possible modification sites, the N-terminal M_1 and K_7. Upon ETD of the singly modified peptide a single loss of 93 Da from both the singly and doubly charged reduced species reflects the incorporation of a single NN modification (**Figure 7.3A**), whereas the doubly modified peptide shows two sequential losses of 93 Da, thus indicating two modifications (**Figure 7.3 B**).

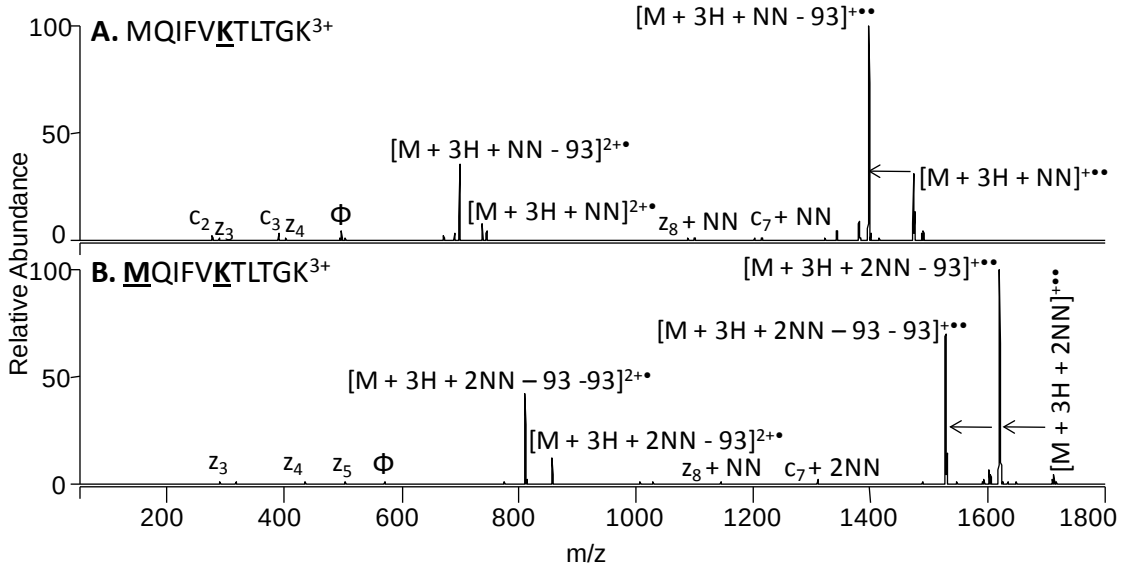

Figure 7.3 Preferential ETD cleavage of NN-modified peptides. ETD of triply charged MQIFVKTLTGK with a single NN modification (A) and double NN modification (B). The sites of modification are indicated by bold underlined font. Preferential ETD cleavage is indicated by the arrows. Φ indicates the precursor ion in each spectrum.

In this example, the sites of modification can be pinpointed directly from the ETD spectra. For the singly modified peptide, the unmodified c_2, c_3, z_3, and z_4 ions and the modified z_8 and c_7 ions localize the first modification on K_7; for the doubly modified the absence of the unmodified c_2, c_3, z_3, z_4, z_5, singly modified z_8, and the doubly modified c_7 ions indicates the second modification is located at M_1. However, in some cases the location of modification cannot be pinpointed from the ETD spectrum alone due to insufficient fragment ions. Thus, a subsequent CID (MS^3) step was included to target the peptide after loss of 93 Da from the charge reduced species in a data dependent manner (see **Figure 7.4** for an example from ubiquitin). This additional step ensured the ability to sequence each modified peptide.

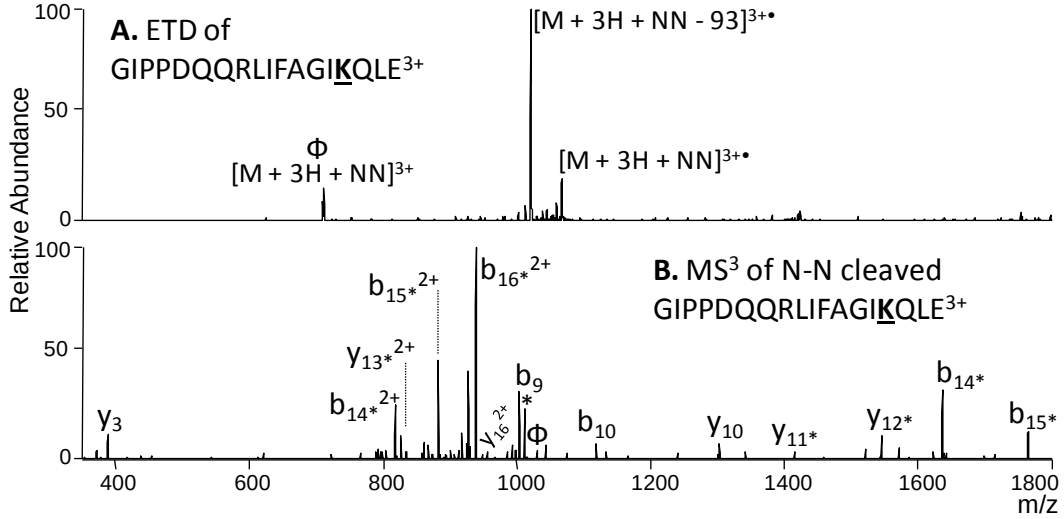

Figure 7.4 Identification of NN-modified peptides from ubiquitin; A. ETD of triply-charged, singly modified GIPPDQQRLIFAGIKQLE, B. CID of NN-cleaved peptide following ETD (precursor m/z 1021). An ETD reaction time of 100 msec and a collision energy of 35% was used. The site of modification is indicated by bold underlined font. * indicates water/ammonia losses. Φ indicates the precursor ion.

For ubiquitin the most accessible residues are K_6, K_{11}, K_{48} and K_{63} all with calculated solvent accessibilities over 50% as shown in **Table 7.1**. For all three digests K_6, K_{48}, K_{11} and K_{63} are among the most accessible sites with the best agreement obtained for the GluC and ArgC digests. K_{27} and K_{29} were shown to be the least accessible sites as they displayed no reactivity based on the GluC and ArgC results and very little reactivity for the set of trypsin results. These sites are heavily involved in hydrogen bonding on the interior of the protein. Our NN-based accessibility results are in general agreement with other surface studies that have been conducted for ubiquitin.[43, 44] The ArgC results gave the best correlation, revealing that K_{11} and K_{63} were the most accessible free amines while K_6 and

K_{48} were slightly less accessible but much more so than the other four possible modification sites which were unreactive. The data obtained for the ArgC digest did not indicate any modification of site K_{33} which was surprising since the ESI-MS/MS results for the other two digests showed high reactivity for this residue. This may be due to the large size of the peptide that is anticipated for incorporation of the modification at this residue, thus making it unlikely to being detected.

Our results also agree well with previous surface accessibility experiments performed on ubiquitin. Upon amidination of the intact protein, the Reilly group found that the only inaccessible Lys was K_{27} while all other possible modification sites were fully amidinated.[43] This Lys residue resides in the bottom of a hydrophobic pocket of ubiquitin's tertiary structure, thus leaving it completely shielded from the solvent. For all three digests, using our NN reagent, Lys27 was observed with the lowest reactivity having only 1% reactivity based on results for the tryptic digest and 0% reactivity based on the analysis of both the GluC and ArgC digests. A top-down approach that utilized N-hydroxysuccinimidyl acetate to acetylate primary amines in ubiquitin found the reactivity trend to be Met1 ≈ Lys6 ≈ Lys48 ≈ Lys 63 > Lys 33 > Lys 11> Lys 27, Lys 29.[44] The Met1, Lys6, Lys48 and Lys63 sites were found to be the most reactive as they were involved in only weak hydrogen bonds involving other backbone carbonyl groups, while Lys 11, Lys 27, and Lys 29 were the least accessible due to their involvement in strong hydrogen bonds to carboxylic acids on other amino acid side chains. Our general reactivity trend was Lys6 ≈ Lys63 ≈ Lys 48 > Met 1 > Lys 33 > Lys 11 > Lys 27, Lys 29; in which the only discrepancy was for the N-terminal methionine which was

found to be slightly less reactive than Lys 6, Lys 63, and Lys 48. This trend agrees well with the solvent accessibility predictions for the tertiary structure of ubiquitin from GetArea as the N-terminus had only a 14% surface accessibility while Lys 6, Lys 63 and Lys 48 all averaged about 60% reactivities.

7.4.3 WHEAT EIF4E

Upon successful demonstration of ETD-selective cleavage of the NN-modified peptides, the surface maps of two other proteins were evaluated using a similar strategy. The first, eukaryotic translation initiation factor-4E (eIF4E), is a crucial protein for the initiation of protein synthesis.[45] The structure has been extensively studied in mammalian and yeast cells but only recently in wheat. Upon crystallization, a dimeric structure has been observed due to the artifactual formation of a disulfide bond. The only known monomeric structure was determined for a mutant bound to 7-methyl-GDP that cannot form the disulfide bond.[45] Wheat eIF4E has 15 Lys and the N-terminus as possible modification sites. At the low molar ratios utilized for ubiquitin, minimal modifications were observed; however, at higher protein/NN molar ratios of 1:15 to 1:20, the dominant species was the singly modified protein (**Figure 7.5A**).

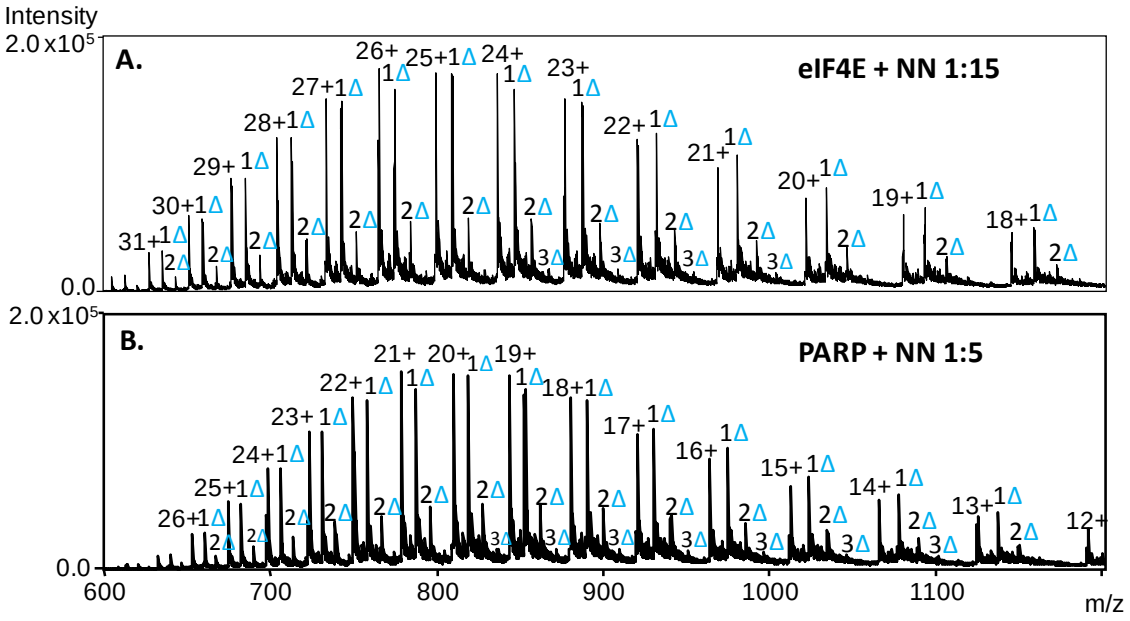

Figure 7.5 ESI-mass spectra of the NN-modified proteins. **A.** eIF4E at a 1:15 protein:NN molar ratio, **B**. PARP at a 1:5 protein:NN molar ratio; **Δ**= NN adduct

The relatively high protein/NN ratio required for this protein, as well as the rather low modification rate, suggests that there are fewer highly accessible Lys residues and the native structure is only minimally disrupted after the first modification. The NN-modified protein was then subjected to enzymatic digestion using three proteases (trypsin, GluC, ArgC), followed by LCMS analysis. In each case, spectral acquisition involved collection of full ESI mass spectra, followed by ETD spectra of the five most abundant ions and CID spectra of the most abundant fragment ion in each ETD spectrum. All peptides were manually identified. The characteristic loss of 93 Da in the ETD spectrum was used to pinpoint the NN-modified peptides, and the subsequent CID spectrum was used to identify the sequence (see **Figure 7.6** for an example from eIF4E).

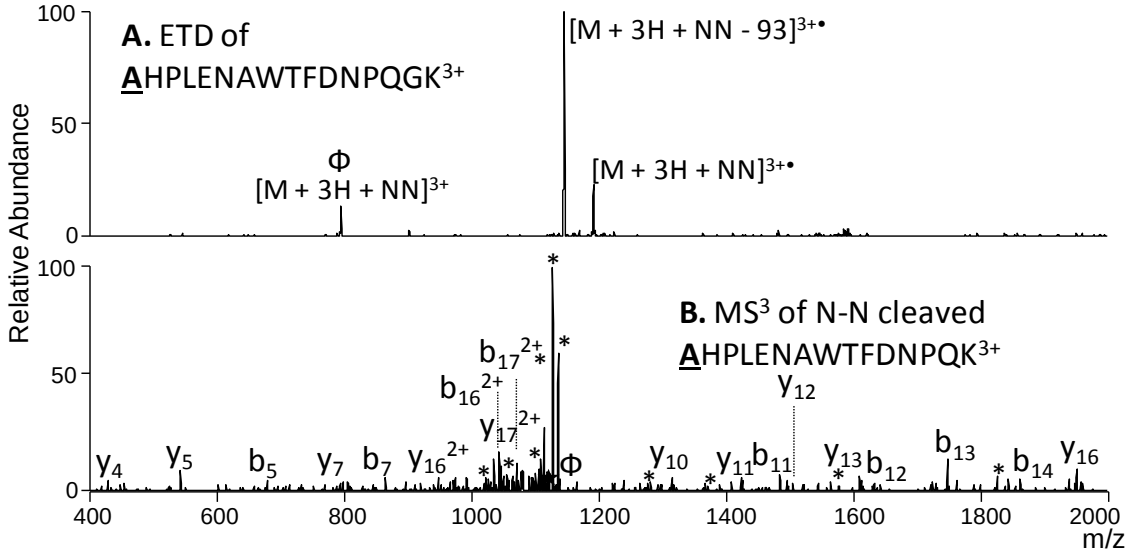

Figure 7.6 Identification of NN-modified peptides from wheat eIF4E; A. ETD of triply-charged, singly modified AHPLENAWTFWFDNPQGK, B. CID of NN-cleaved peptide following ETD (precursor m/z 1145). ETD reaction time of 100 msec and collision energy of 35% was used. The site of modification is indicated by bold underlined font. * indicates water/ammonia losses. Φ indicates precursor ion.

Unmodified peptides were identified using the initial ETD spectrum. The peak area for each modified residue was integrated and compared to all peptides containing the unmodified residue using QualBrowser software.

The artifactual dimeric structure of wheat eIF4E determined previously by X-ray and NMR measurements that arises from disulfide bond formation does not give a good representation of the monomer due to the extensive protein-protein interface. A structure for a known mutant of wheat eIF4E has also been characterized previously as well, but this mutant is formed upon binding to 7-methyl-GDP which may cause considerable changes in tertiary structure, especially near the binding pocket. Therefore the tertiary structure of monomeric eIF4E was also predicted using the ITASSER program prior to calculation of solvent accessibility values for all amino acid side-chains via the GetArea

program and compared to the known mutant monomer and dimeric forms (**Table 7.2**). The known structure for the dimer, the 7-methyl-GDP mutant, and the monomer structure predicted from ITASSER are shown in **Figure 7.7** with key Lys residues labeled.

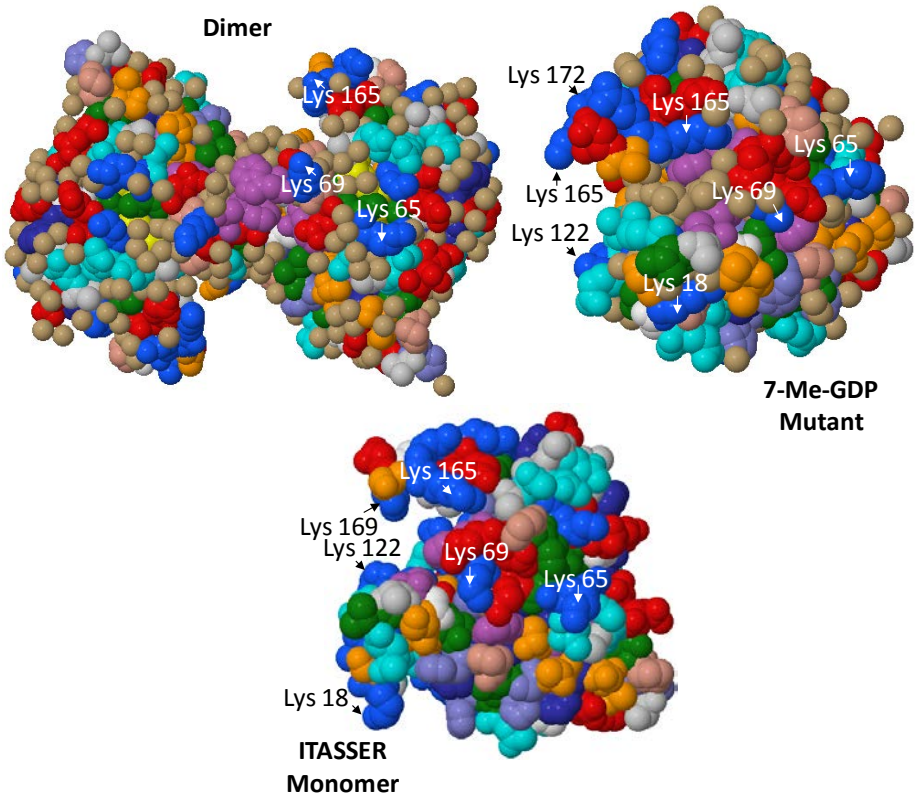

Figure 7.7 Structures for wheat eIF4E; dimer = PDB 2IDR, 7-Me-GDP mutant = PDB 2IDV. The monomer is a predicted model from ITASSER, and key Lys residues (blue) are labeled.

The solvent accessibility values calculated by GetArea for each structure were compared to the experimental percent reactivities determined for each enzymatic digest in **Table 7.2**.

Residue	Solvent Accessibility %			% Reactivity		
	Dimer	Mutant	Monomer	Tryptic	GluC	ArgC
Ala 1	46	65	41	2	11	5
Lys 18	43	34	96	6	5	47
Lys 52	76	74	50	0	0	0
Lys 63	5	9	20	0	NA	0
Lys 65	70	100	94	22	NA	11
Lys 69	100	20	48	0	0	0
Lys 80	27	20	20	0	NA	0
Lys 89	77	80	67	NA	NA	NA
Lys 122	40	38	92	61	5	70
Lys 131	31	26	31	0	0	0
Lys 144	68	73	66	0	2	0
Lys 147	22	15	13	0	0	0
Lys 153	42	42	84	0	0	0
Lys 165	80	43	50	0	0	0
Lys 169	40	81	91	45	9	34
Lys 172	28	94	72	0	0	0

Table 7.2 Comparison of Solvent Accessibility to Percent Reactivity for wheat eIF4E reacted with NN at a 1:15 protein:NN molar ratio. Solvent accessibility was calculated using GetArea. Percent Reactivity was calculated using equation 7.1. “NA” indicates a peptide that was not found in neither its modified or unmodified form. Standard deviation ranged from 10-20% for all reacted sites.

Based on the solvent accessibilities calculated for the monomer-predicted structure, three to four Lys residues are expected to exhibit much higher accessibility (above 90%) than

all other possible modification sites; K_{18}, K_{65}, K_{122}, and K_{169}. For the ESI-MS/MS results of the tryptic digest of the NN-modified protein, all four of these sites were identified as the primary sites of modification, thus indicating substantial NN reactivities. The N-terminus (A_1) exhibited low accessibility (2%). Upon increasing the molar ratio of protein:NN from 1:15 to 1:20, the reactivity for K_{65}, K_{122}, and K_{169} all increased to 100% whereas the reactivity of K_{18} showed little change and that of the N-terminus increased to 26%. The latter change may signal a structural change for this part of the protein upon binding of NN to K_{18}, thus opening up the structure so that the N-terminus is more readily accessible. For the GluC digest, significant reactivity was observed for A_1, K_{18}, K_{122}, and K_{169} as seen in the tryptic digest. However, based on the GluC digest, no reactivity values could be determined for K_{63}, K_{65}, K_{80} or K_{89} as no peptides were identified that contained these residues. We believe the absence of these peptides arises from the size of the digested peptide produced and possible spontaneous disulfide bonds forming following digestion. For the ArgC digest, again K_{18}, K_{65}, K_{122} and K_{169} proved to be the most reactive residues with A_1 showing a small amount of reactivity at a 1:15 molar ratio. The lack of reactivity of other Lys sites indicates these Lys residues reside on the interior of the protein leaving them unable to react with the NN reagent.

Based on the solvent accessibilities calculated for the ITASSER-predicted monomer, three to four Lys residues are expected to exhibit much higher accessibility (above 90%) than all other possible modification sites; K_{18}, K_{65}, K_{122}, and K_{169}. For the ESI-MS/MS results of the tryptic digest of the NN-modified protein, all four of these sites were identified as the primary sites of modification, thus indicating substantial NN reactivities. The N-terminus (A_1) exhibited low accessibility (2%). Upon increasing the molar ratio of protein:NN from

1:15 to 1:20, the reactivity for K_{65}, K_{122}, and K_{169} all increased to 100% whereas the reactivity of K_{18} showed little change and that of the N-terminus increased to 26%. The latter change may signal a structural change for this part of the protein upon binding of NN to K_{18}, thus opening up the structure so that the N-terminus is more readily accessible. For the GluC digest, significant reactivity was observed for A_1, K_{18}, K_{122}, and K_{169} as seen in the tryptic digest. However, based on the GluC digest, no reactivity values could be determined for K_{63}, K_{65}, K_{80} or K_{89} as no peptides were identified that contained these residues. We believe the absence of these peptides arises from the size of the digested peptide produced and possible spontaneous disulfide bonds forming following digestion. For the ArgC digest, again K_{18}, K_{65}, K_{122} and K_{169} proved to be the most reactive residues with A_1 showing a small amount of reactivity at a 1:15 molar ratio. The lack of reactivity of other Lys sites indicates these Lys residues reside on the interior of the protein leaving them unable to react with the NN reagent.

In comparing the experimental percent reactivities to the solvent accessibilities of the dimer, mutant, and TASSER-predicted monomer (**Table 7.2**), a few key variations are noted. In both the dimer and mutant forms, K_{18} and K_{122} are much less accessible whereas in the predicted monomer form, these residues have solvent accessibilities of 96% and 92% respectively, making them two of the most accessible sites. This notable difference mirrors the percent reactivities obtained from the NN protein modification results, confirming that both of these sites are two of the most reactive in wheat eIF4E. K_{69} is 100% accessible in the dimer but is completely unreactive with NN based on analysis of the various protein digests, indicating it resides in the interior of the monomer. K_{172} for the 7-Me-GDP mutant is expected to be located on the protein surface as a highly accessible residue, yet the

experimental percent reactivity value is zero. The site of this amino acid coincides with the binding location of the ligand 7-methyl-GDP in the mutant, thus suggesting that the bound ligand induces a significant structural change that does not occur in the native monomer. In light of these differences in predicted accessibilities and experimentally-determined reactivities, the NN reactivity results support the ITASSER predicted structure for the native monomer.

7.4.4 PARP-1 DOMAIN C

Poly(ADP-ribose) polymerase-1 (PARP-1) is a multimodular protein consisting of six domains (domains A-F) that is involved in several key biological processes including DNA repair and cell death.[46, 47] Domain C has been shown to be vital to these functions and without it, PARP-1 ABDEF exhibits no activity. Domain C, by itself, also does not bind to DNA, confirming that it is the interactions of the entire multi-domain protein that stimulates biological activity. Domain C has recently been isolated and studied by NMR to determine its solution structure. However, upon expression, only a shortened version with three fewer amino acids at its N-terminus and 15 fewer amino acids on its C-terminus was identified by NMR spectroscopy, and no solution structural data exists for the true full form of the protein in its monomer form. In its full form, domain C contains 18 Lys and the N-terminus as possible modification sites whereas the truncated version maintains all 18 Lys residues but with K_4 is also the N-terminus. Upon reaction of PARP with NN, several residues are modified even at low molar ratios (**Figure 7.5B**). The NN-modified protein was enzymatically digested using trypsin, GluC, or ArgC, and the resulting peptides were analyzed by the $LCMS^n$ strategy described above and

demonstrated in **Figure 7.8** (data dependent analysis entailing ETD followed by CID of the most abundant products).

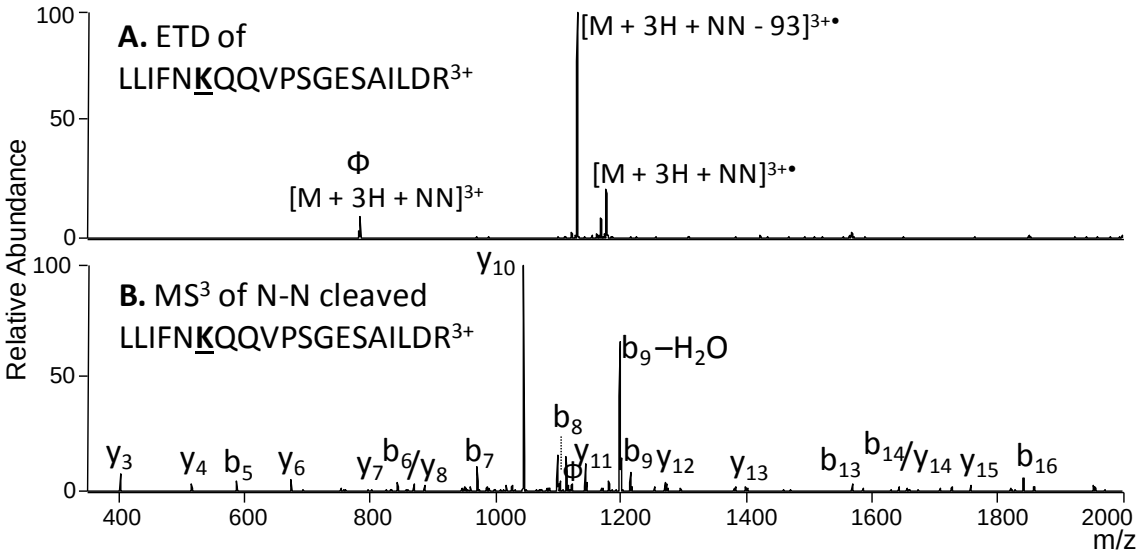

Figure 7.8 Identification of NN-modified peptides from PARP-1 Domain C. A. ETD of triply-charged, singly modified LLIFNKQQVPSGESAILDR, B. CID of NN-cleaved peptide following ETD (precursor m/z 1130). An ETD reaction time of 100 msec and a collision energy of 35% was used. The site of modification is indicated by bold underlined font. Φ indicates the precursor ion in each spectrum.

Percent reactivities were calculated and summarized in **Table 7.3**.

Residue	Surface Accessibility %						% Reactivity		
	Shortened	IT1	IT2	IT3	IT4	IT5	Tryptic	GluC	ArgC
GLY 1	--	76	90	100	72	79	98	100	100
LYS 4	100	87	78	91	76	88	98	85	95
LYS 7	81.6	70	79	71	75	91	41	38	5
LYS 10	85.9	76	63	64	71	77	25	0	10
LYS 20	20.6	16	20	29	30	10	14	0	9
LYS 24	65.2	56	48	37	70	40	7	0	0
LYS 25	65.4	55	62	57	58	56	0	0	0
LYS 33	49.6	35	33	34	34	56	0	0	0
LYS 40	93.1	76	87	85	77	82	4	1	8
LYS 76	94.9	62	45	45	48	62	0	0	7
LYS 91	54	67	55	63	67	45	0	0	0
LYS 95	73.7	50	36	42	41	44	0	0	0
LYS 102	64.8	69	66	63	52	37	52	0	0
LYS 108	70.3	67	92	84	74	71	23	65	35
LYS 117	25.3	37	42	42	37	13	16	0	0
LYS 118	88.3	46	50	46	55	62	0	100	98
LYS 120	24.6	41	53	50	52	26	37	0	0
LYS 122	43.4	84	83	74	76	53	0	0	0
LYS 123	100	84	83	84	80	54	43	0	0

Table 7.3 Comparison of Surface Accessibility to Percent Reactivity for PARP-1 Domain C reacted with NN in a 1:5 molar ratio protein:NN. Solvent accessibility for shortened structure was calculated using GetArea for pdb 2JVN and for the full native structure based on the top five structures predicted by ITASSER (IT1-IT5). Percent reactivity was calculated using equation 7.1. The bold box highlights the best model structure correlating with percent reactivity. Standard deviations ranged from 10-20% for all reacted sites.

Since the tertiary structure of the full sequence version of domain C was unknown, its conformation was predicted by ITASSER,[1] resulting in five structures whose surface accessibilities were predicted by the GetArea program. The surface accessibilities for the shortened domain C structures are shown in **Table 7.3**. The percent reactivities obtained for the NN-modified peptides do not parallel the surface accessibilities predicted from the NMR structure, as reflected by the differences in reactivities of a few key residues, including the highly reactive N-terminal G_1 and the non-reactive K_{123} which was predicted to be 100% accessible based on the NMR structure. Of the structures predicted by ITASSER, the fifth model (IT5) showed the best correlation with the experimental surface accessibility results. This structure showed a highly reactive N-terminal portion of the protein with G_1, K_4, K_7, and K_{10} all having solvent accessibility values above 75% while the K_{123} showed somewhat lower accessibility at 54%. Upon examination of the structure of IT5, the low reactivity of K_{123} may be rationalized by steric blocking by the chain of amino acids on the adjacent C-terminus (**Figure 7.9**).

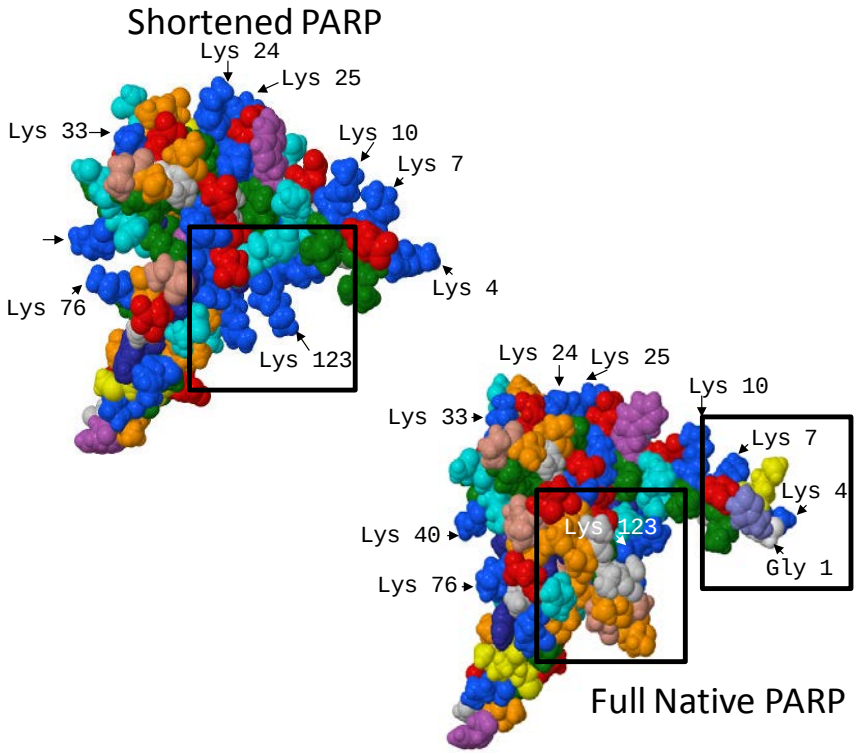

Figure 7.9 Structures for PARP-1 Domain C, including shortened PARP = PDB 2JVN and the full native PARP is predicted model 5 from ITASSER. Lys residues (blue) with high accessibility are labeled. Boxed sections show key changes in structure recognized based on NN reactivity.

The percent reactivities and predicted surface accessibilities of the other regions of the protein show good agreement with K_{40}, K_{76}, K_{108}, and K_{118} all predicted to be more than 60% accessible, in line with the values from the NN modification reactions. The only residue that does not correlate between the percent reactivity and solvent accessibility is K_7. It was predicted as one of the more accessible residues but in all cases was shown to have poor reactivity in comparison to G_1 and K_4. We hypothesize that this deviation is due to structural effects caused upon modification. Since G_1 and K_4 are so highly reactive, we

suspect that upon their modification by NN, the K_7 site is blocked, therefore hindering its reaction with the NN reagent.

7.5 CONCLUSIONS

A lysine-reactive surface accessibility reagent that incorporates an ETD-selectively cleavable N-N hydrazone bond is reported. Enzymatic digestion of the surface modified proteins results in an array of peptides, some of which contain the hydrazone moiety. Upon ETD, the hydrazone bond cleaves selectively and with high efficiency to yield a loss of 93 Da from the charge-reduced peptide. CID of the resulting neutral loss products via a data dependent MS3 strategy allows sequencing of the peptides and the modification sites to be pinpointed. This selective ETD/CID method is demonstrated for several proteins and provides maps of surface accessibility that are consistent with the predicted structures of the proteins.

7.6 REFERENCES

(1) Roy, A.; Kucukural, A.; Zhang, Y. *Nat. Protoc.* **2010**, *5*, 725-738.

(2) Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785-815.

(3) Jorgensen, T. D.; Gardsvoll, H.; Ploug, M.; Roepstorff, P. *J. Am. Chem. Soc.* **2005**, *127*, 2785-2793.

(4) Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R. *Anal. Chem.* **2009**, *81*, 7892-7899.

(5) Rand, K. D.; Adams, C. M.; Zubarev, R. A.; Joergensen, T. J. D. *J. Am. Chem. Soc.* **2008**, *130*, 1341-1349.

(6) Rand, K. D.; Zehl, M.; Jensen, O. N.; Jorgensen, T. J. D. *Anal. Chem.* **2009**, *81*, 5577-5584.

(7) Petrotchenko, E. V.; Borchers, C. H. *Mass Spectrom. Rev.* **2010**, *29*, 862-876.

(8) Suckau, D.; Mak, M.; Przybylski, M. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5630-5634.

(9) Glocker, M. O.; Borchers, C.; Fiedler, W.; Suckau, D.; Przybylski, M. *Bioconjugate Chem.* **1994**, *5*, 583-590.

(10) Zappacosta, F.; Ingallinella, P.; Scaloni, A.; Pessi, A.; Bianchi, E.; Sollazzo, M.; Tramontano, A.; Marino, G.; Pucci, P. *Protein Sci.* **1997**, *6*, 1901-1909.

(11) Izumi, S.; Kaneko, H.; Yamazaki, T.; Hirata, T.; Kominami, S. *Biochemistry* **2003**, *42*, 14663-14669.

(12) Turner, B. T., Jr.; Sabo, T. M.; Wilding, D.; Maurer, M. C. *Biochemistry* **2004**, *43*, 9755-9765.

(13) Scholten, A.; Visser, N. F. C.; van den Heuvel, R. H. H.; Heck, A. J. R. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 983-994.

(14) Liu, X.; Reilly, J. P. *J. Proteome Res.* **2009**, *8*, 4466-4478.

(15) Running, W. E.; Reilly, J. P. *Proteomics* **2010**, *10*, 3669-3687.

(16) Lehman, J. A.; Hoelz, D. J.; Turchi, J. J. *Biochemistry* **2008**, *47*, 4359-4368.

(17) Shell, S. M.; Hess, S.; Kvaratskhelia, M.; Zou, Y. *Biochemistry* **2005**, *44*, 971-978.

(18) Sharp, J. S.; Nelson, S.; Brown, D.; Tomer, K. B. *Virology* **2006**, *348*, 216-223.

(19) Blodgett, D. M.; De Zutter, J. K.; Levine, K. B.; Karim, P.; Carruthers, A. *J. Gen. Physiol.* **2007**, *130*, 157-168.

(20) Azim-Zadeh, O.; Hillebrecht, A.; Linne, U.; Marahiel, M. A.; Klebe, G.; Lingelbach, K.; Nyalwidhe, J. *J. Biol. Chem.* **2007**, *282*, 21609-21617.

(21) Vasilescu, J.; Guo, X.; Kast, J. *Proteomics* **2004**, *4*, 3845-3854.

(22) Kosower, E. M.; Kosower, N. S. *Methods Enzymol.* **1995**, *251*, 133-148.

(23) Alley, S. C.; Ishmael, F. T.; Jones, A. D.; Benkovic, S. J. *J. Am. Chem. Soc.* **2000**, *122*, 6126-6127.

(24) Sinz, A.; Kalkhof, S.; Ihling, C. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1921-1931.

(25) Hurst, G. B.; Lankford, T. K.; Kennel, S. J. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 832-839.

(26) Fujii, N.; Jacobsen, R. B.; Wood, N. L.; Schoeniger, J. S.; Guy, R. K. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 427-429.

(27) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80*, 4807-4819.

(28) Pearson, K. M.; Pannell, L. K.; Fales, H. M. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 149-159.

(29) Mueller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem.* **2001**, *73*, 1927-1934.

(30) Collins, C. J.; Schilling, B.; Young, M.; Dollinger, G.; Guy, R. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4023-4026.

(31) Huang, B. X.; Kim, H.-Y.; Dass, C. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1237-1247.

(32) Taverner, T.; Hall, N. E.; O'Hair, R. A. J.; Simpson, R. J. *J. Biol. Chem.* **2002**, *277*, 46487-46492.

(33) Davidson, W. S.; Hilliard, G. M. *J. Biol. Chem.* **2003**, *278*, 27199-27207.

(34) Bennett, K. L.; Kussmann, M.; Bjork, P.; Godzwon, M.; Mikkelsen, M.; Sorensen, P.; Roepstorff, P. *Protein Sci.* **2000**, *9*, 1503-1518.

(35) Back, J. W.; Hartog, A. F.; Dekker, H. L.; Muijsers, A. O.; de Koning, L. J.; de Jong, L. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 222-227.

(36) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Anal. Chem.* **2005**, *77*, 311-318.

(37) Gardner, M. W.; Brodbelt, J. S. *Anal. Chem.* **2010**, *82*, 5751-5759.

(38) Reid, G. E.; Roberts, K. D.; Simpson, R. J.; O'Hair, R. A. J. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1131-1150.

(39) Roberts, K. D.; Reid, G. E. *J. Mass Spectrom.* **2007**, *42*, 187-198.

(40) Zhou, X.; Lu, Y.; Wang, W.; Borhan, B.; Reid, G. E. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1339-1351.

(41) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. *Proteomics* **2011**, *11*, 1329-1334.

(42) Fraczkiewicz, R.; Braun, W. *J. Comput. Chem.* **1998**, *19*, 319-333.

(43) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2005**, *77*, 7274-7281.

(44) Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. *Journal of Mass Spectrometry* **2004**, *39*, 322-328.

(45) Monzingo, A. F.; Dhaliwal, S.; Dutt-Chaudhuri, A.; Lyon, A.; Sadow, J. H.; Hoffman, D. W.; Robertus, J. D.; Browning, K. S. *Plant Physiol.* **2007**, *143*, 1504-1518.

(46) Satoh, M. S.; Lindahl, T. *Nature* **1992**, *356*, 356-358.

(47) Yu, S.-W.; Wang, H.; Poitras, M. F.; Coombs, C.; Bowers, W. J.; Federoff, H. J.; Poirier, G. G.; Dawson, T. M.; Dawson, V. L. *Science* **2002**, *297*, 259-263.

Chapter 8

Conclusions

Mass spectrometry has proven extremely fruitful in the characterization and identification of biomolecules due to its sensitivity, speed, and ability to delve into their structure through fragmentation. As the field continues to progress and samples become more complex, the ability to distinguish between analytes and examine constituents of both high and low abundances becomes increasingly important. Selective derivatization can provide constraints on which analytes are targeted and identify specific structural details of interest. Tandem mass spectrometry remains an essential tool for creating diagnostic fingerprints of biological molecules and is the most common methodology for protein identification. However, most MS/MS methods yield partial sequence coverage, averaging around 75%, which is not always sufficient to assemble the complete primary sequence of a newly discovered protein or map its modifications. When tandem mass spectrometry is coupled with selective derivatization methods, advantageous chemistry can be used to enhance the mass spectrometric analysis. Several methods of incorporating a selective tag into proteins to enhance tandem mass spectrometry were discussed in this dissertation.

In Chapter 3, a phosphate chromophore was added to the N-terminus of digested peptides to promote dissociation following IR irradiation. The efficiency of this derivatization method coupled to IRMPD was compared to traditional CID for sequencing of peptides to highlight the benefits of IRMPD, including the elimination of the low mass cut-off and increased sequence coverage due to secondary dissociation of

uninformative neutral losses from the precursor ion. The negative group of the phosphate also served to limit the fragmentation so that a sole series of ions was detected in the mass spectrometry, the *y*-series. The greater sequence coverage combined with simplified fragmentation spectra provided by IRMPD promoted greater confidence in the sequencing and identification of peptides and proteins. A novel phosphate isothiocyanate reagent was successfully synthesized and used to demonstrate enhanced photodissociation efficiency for a tryptic digest.

High resolution is a key component to the analysis of large intact biomolecules. Chapter 4 discussed the integration of an IR laser with an orbitrap mass spectrometer to extend the use of IRMPD analysis to large intact biological molecules with high resolution/high mass accuracy detection. With the growing popularity of orbitrap mass analyzers, the introduction of PD on this type of instrument offers an attractive alternative to conventional CID. PD was conducted in the higher energy collision (HCD) cell where ions could be trapped and irradiated before being transmitted through the C-trap for orbitrap detection. IRMPD provided an equal or more diverse set of fragment ions compared to higher energy CID while also offering selectivity for chromogenic analytes.

Vacuum ultra-violet photodissociation is a higher energy single photon method to fragment peptides and direct fragmentation. The high energy deposition of UVPD leads to an extremely diverse set of fragment ions that can provide greater sequence coverage as well as side chain cleavages. While it has been found that UVPD in the vacuum range is possible for peptides, the sensitivity of the method depends on the photoabsorption cross-sections of the peptides, with those peptides that contain aromatic amino acids exhibiting higher UVPD efficiencies. Chapter 5 discussed the use of two aromatic

derivatization reagents that were used to enhance the cross-sections of peptides to increase their photodissociation effiencies at 193 nm and increase the sensitivity of the method. Attachment of aromatic groups to tryptic peptides of BSA illustrated the increase in dissociation efficiencies and sequence coverages upon UVPD.

Electron transfer dissociation has proven to be a complementary technique to collision induced dissociation for the analysis of peptides/proteins. However, ions in low charge states predominantly undergo charge reduction without dissociation, thus limiting the ability to effectively sequence these peptides. In chapter 6, the selective derivatization of cysteines with a quaternary amine (i.e. fixed charge) reagent shifted the charge states to higher values by the number of cysteines present in each peptide. The increase in the number of charges destabilized the charge-reduced species leading to more prominent backbone fragments. In addition, upon ETD a reporter ion was created, thus allowing differentiation of cysteine-containing peptides from other peptides in a complex mixture. The addition of a fixed charge also increased the ionization efficiencies of derivatized peptides, leading to greater detection sensitivity. Three cysteine-alkylation methods were compared for a tryptic digest to prove the analytical merits of the fixed charge alkylation method.

Lastly, to demonstrate the ability to use mass spectrometry to gain information about a protein's tertiary structure, chapter 7 discusses a novel derivatization of surface accessible amines that upon ETD activation leads to a characteristic fragmentation pattern that can be used to quickly identify surface residues. Surface mapping with mass spectrometry has been extremely important due to the large number of proteins that prove difficult to crystallize or purify in quantities insufficient for NMR analysis. While mass

spectrometry only affords low resolution structural analysis, the sample consumption is minimal. In addition, with the addition of a preferentially cleavable bond, like a hydrazone for ETD, the characteristic fragmentation spectrum created for all modified residues eliminated the needle-in-a-haystack data analysis problem that arises when trying to identify modified peptides in complex mixtures.

In summary, CID, IRMPD, UVPD and ETD were utilized to characterize proteins derivatization with either a chromophore or a charged moiety. These derivatization methods added selectivity to the analytical workflow, ultimately leading to improvements in peptide sequence coverage and the ability to target specific types of peptides in complex mixtures.

Future work will focus on utilizing these methods on more complex samples such as cell lysates or protein complexes to demonstrate their application in more complex proteomic applications. For example, the protein PARP 1 Domain C from chapter 7 is a single part of a six part complex and it is only the entire protein complex that proves active. PARP-1 ABDEF showed no activity while domain C by itself also remained inactive showing that it is the interactions of the entire multi-domain protein that stimulate biological activity. Therefore investigating any structural changes upon incorporation of Domain C into the macromolecular protein complex would allow mapping of the protein-protein interactions. Moreover, the development of other reagents that target different amino acid residues would expand the range of applications possible. For example, instead of affixing a charge at all cysteine residues, other amino acids that are more or less abundant in proteins could be targeted. Attachment of fixed charge sites could well change the fragmentation patterns of the resulting peptides too and perhaps

lead to other trackable reporter ions. In addition, thus far all the work discussed in this dissertation was conducted in the positive ionization mode. However, nearly half the proteome is acidic, leaving many avenues to be explored that would increase the ionization and dissociation of peptide anions. Many new dissociation techniques have been developed to improve peptide anion sequencing but many suffer from low efficiency or low sequence value, both of which might be overcome through the use of selective derivatization.

Vita

Lisa Anne Vasicek was born in Tulsa, Oklahoma in February 1984 to Dr. Daniel Vasicek and Clarice Vasicek. After earning an International Baccalaureate degree from Booker T. Washington High School in spring of 2002, she continued her education at the University of Tulsa in the fall of 2002. She graduated cum laude with a bachelor's of science degree in chemistry in May 2006. While working towards her undergraduate degree, she began research on the characterization of derivatized quantum dots using capillary electrophoresis under the guidance of Professor Kenneth P. Roberts. Following graduation, she moved to Austin, TX to begin her graduate studies the University of Texas at Austin under the direction of Professor Jennifer S. Brodbelt. Her current work focuses on developing derivatization techniques to promote selective dissociation patterns using mass spectrometry that can be used to improve identification and characterization of proteins for proteomic applications.

Permanent email: la.vasicek@gmail.com

This dissertation was typed by the author.